Submitted to the IEEE Transactions on Information Theory, October 2006.

On the Role of Estimate-and-Forward with Time-Sharing in Cooperative Communication

Ron Dabora Sergio D. Servetto

Abstract

In this work we focus on the general relay channel. We investigate the application of estimate-and-forward (EAF) to different scenarios. Specifically, we consider assignments of the auxiliary random variables that always satisfy the feasibility constraints. We first consider the multiple relay channel and obtain an achievable rate without decoding at the relays. We demonstrate the benefits of this result via an explicit discrete memoryless multiple relay scenario where multi-relay EAF is superior to multi-relay decode-and-forward (DAF). We then consider the Gaussian relay channel with coded modulation, where we show that a three-level quantization outperforms the Gaussian quantization commonly used to evaluate the achievable rates in this scenario. Finally we consider the cooperative general broadcast scenario with a multi-step conference. We apply estimate-and-forward to obtain a general multi-step achievable rate region. We then give an explicit assignment of the auxiliary random variables, and use this result to obtain an explicit expression for the single common message broadcast scenario with a two-step conference.

I. Introduction

The relay channel was introduced by van der Meulen in 1971 [1]. In this setup, a single transmitter with channel input X^n communicates with a single receiver with channel output Y^n, where the superscript n denotes the length of a vector. In addition, an external transceiver, called a relay, listens to the channel and is able to output signals to the channel. We denote the relay output with Y_1^n and its input with X_1^n. This setup is depicted in figure 1.

A. Relaying Strategies

In [2] Cover & El-Gamal introduced two relaying strategies commonly referred to as decode-and-forward (DAF) and estimate-and-forward (EAF). In DAF the relay decodes the message sent from the transmitter and then, at the next time interval, transmits a codeword based on the decoded message. The rate achievable with DAF is given in [2, theorem 1]:

Theorem 1: (achievability of [2, theorem 1]) For the general relay channel any rate R satisfying

$$R \leq \min\{I(X, X_1; Y), I(X; Y_1|X_1)\} \tag{1}$$

The authors are with the School of Electrical and Computer Engineering, Cornell University, Ithaca, NY. URL: `http://cn.ece.cornell.edu/`. Work supported by the National Science Foundation, under awards CCR-0238271 (CAREER), CCR-0330059, and ANR-0325556.

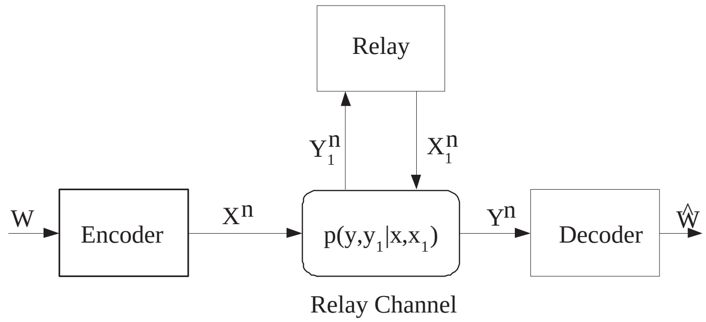

Fig. 1. The relay channel. The encoder sends a message W to the decoder.

for some joint distribution $p(x, x_1, y, y_1) = p(x, x_1)p(y, y_1|x, x_1)$*, is achievable.*

We note that for DAF to be effective, the rate to the relay has to be greater than the point-to-point rate i.e.

$$I(X; Y_1|X_1) > I(X; Y|X_1), \tag{2}$$

otherwise higher rates could be obtained without using the relay at all. For relay channels where DAF is not useful or not optimal, [2] proposed the EAF strategy. In this strategy, the relay sends an estimate of its channel input to the destination, without decoding the source message at all. The achievable rate with EAF is given in [2, theorem 6]:

Theorem 2: ([2, theorem 6]) For the general relay channel any rate R *satisfying*

$$R \leq I(X; Y, \hat{Y}_1|X_1), \tag{3}$$

$$\textit{subject to } I(X_1; Y) \geq I(Y_1; \hat{Y}_1|X_1, Y), \tag{4}$$

for some joint distribution $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|y_1, x_1)$*, where* $||\hat{\mathcal{Y}}_1|| < \infty$*, is achievable.*

Of course, one can combine the DAF and EAF schemes by performing partial decoding at the relay, thus obtaining higher rates as in [2, theorem 7].

B. Related Work

In recent years, the research in relaying has mainly focused on multiple-level relaying and the MIMO relay channel. In the context of multiple-level relaying based on DAF, several DAF variations were considered. In [3] Cover & El-Gamal's block Markov encoding/succesive decoding DAF method was applied to the multiple-relay case. Later work [4], [5] and [6] applied the so-called regular encoding/sliding-window decoding and the regular encoding/backward decoding techniques to the multiple-relay scenario. In [7] the DAF strategy was applied to the MIMO relay channel. The EAF strategy was also applied to the multiple-relay scenario. The work in [8], for example, considered the EAF strategy for multiple relay scenarios and the Gaussian relay channel, in addition to considering the DAF strategy. Also [9] considered the EAF strategy in the multiple-relay setup. Another approach applied recently to the relay channel is that of iterative decoding. In [10] the three-node network in the half-duplex

regime was considered. In the relay case, [10] uses a feedback scheme where the receiver first uses EAF to send information to the relay and then the relay decodes and uses DAF at the next time interval to help the receiver decode its message. Combinations of EAF and DAF were also considered in [11], where conferencing schemes over orthogonal relay-receiver channels were analyzed and compared. Both [10] and [11] focus on the Gaussian case.

An extension of the relay scenario to a hybrid broadcast/relay system was introduced in [12] in which the authors applied a combination of EAF and DAF strategies to the independent broadcast channel with a single common message, and then extended this strategy to the multi-step conference. In [13] we used both a single-step and a two-step conference with orthogonal conferencing channels in the discrete memoryless framework. A thorough investigation of the broadcast-relay channel was done in [14], where the authors applied the DAF strategy to the case where only one user is helping the other user, and also presented an upper bound for this case. Then, the fully cooperative scenario was analyzed. The authors applied both the DAF and the EAF methods to that case.

C. The Gaussian Relay Channel with Coded Modulation

One important instance of the relay channel we consider in this work is the Gaussian relay channel with coded modulation. This scenario is important in evaluating the rates achievable with practical communication systems, where components in the receive chain, such as equalization for example, require a uniformly distributed finite constellation for optimal operation. In Gaussian relay channel scenarios, most often three types for relaying techniques are encountered:

- The first technique is decode-and-forward. This technique achieves capacity for the physically degraded Gaussian relay channel (see [2, section IV]), and also for more general relay channels under certain conditions (see [11]).
- The second technique is estimate-and-forward, where the auxiliary variable $\hat{Y}_1$ is assigned a Gaussian distribution. For example, in [15, section IV] a Gaussian auxiliary random variable (RV) is used in conjunction with time-sharing at the transmitter, and in [16] the ergodic capacity for full duplex transmission with Gaussian EAF is obtained.
- The third technique is linear relaying, where the relay transmits a weighted sum of all its previously received inputs [15, section V]. An important subclass of this family of relaying functions is when the relay transmits a scaled version of its input. This method is called amplify-and-forward [17], and was later combined with DAF to produce the decode-amplify-and-forward method of [18].

Several recent papers consider the Gaussian relay channel with coded modulation. In [19] the author considered variations of DAF for different practical systems. In [17] DAF and amplify-and-forward were considered for coherent orthogonal BPSK signalling, and in [20] a practical construction that implements a half-duplex EAF coding scheme was proposed.

As indicated by several authors (see [15]) it is not obvious if a Gaussian relay function is indeed optimal. In this paper we show that for the case of coded modulation, there are scenarios where non-Gaussian assignments of the

auxiliary RV result in a higher rate than the commonly applied Gaussian assignment.

D. Main Contributions

In the following we summarize the main contributions of this work:

- We give an intuitive insight into the relay channel in terms of information flow on a graph, and show how to obtain [2, theorem 6] from flow considerations. Using flow considerations we also obtain the rate of the EAF strategy when the receiver uses joint-decoding. A similar expression can be obtained by specializing the result of [22] to the case where the relay does not perform partial decoding. We then show that joint-decoding does not increase the maximum rate of the EAF strategy, and find the time-sharing assignment that obtains the joint-decoding rate from the general EAF expression. We also present another time-sharing assignment that always exceeds the joint-decoding rate.
- We introduce an achievable rate expression for the multiple relay scenario based on EAF, that is also practically computabe. As discussed in section I-A, in the "noisy relay" case EAF outperforms DAF. However, for the multiple relay scenario there is no explicit, computationally practical expression based on EAF that can be compared with the DAF-based result presented in [5], so that the best strategy can be selected. As indicated in [8, remark 22, remark 23], applying general EAF to a network with an arbitrary number of relays is computationally impractical due to the large number of constraints that characterize the feasible region. Therefore, it is interesting to explore a computationally simple assignment that allows to derive a result that extends to an arbitrary number of relays. We also provide an explicit numerical example to demonstrate that indeed there are cases where multi-relay EAF outperforms the multi-relay DAF.
- We consider the optimization of the EAF auxiliary random variable for the Gaussian relay channel with an orthogonal relay. We consider the coded modulation scenario, and show that there are three regions: high SNR on the source-relay link, where DAF is the best strategy, low SNR on the source-relay link in which the common EAF with Gaussian assignment is best, and an intermediate region where EAF with hard-decision per symbol is optimal. For this intermediate SNR region we consider two kinds of hard-decisions: deterministic and probabilistic, and show that each one of them can be superior, depending on the channel conditions.
- Lastly, we consider the cooperative broadcast scenario with a multi-step conference. We present a general rate region, extending the Marton rate region of [21] to the case where the receivers hold a K-cycle conference prior to decoding the messages. We then specialize this result to the single common message case and obtain explicit expressions (without auxiliary RVs) for the two-step conference.

The rest of this paper is organized as follows: in section II we discuss the single relay case. We consider the EAF strategy with time-sharing (TS) and relate it to the EAF rate expression for joint-decoding at the destination receiver. In section III we present an achievable region for the multiple-relay channel, and in section IV we examine the Gaussian relay channel with coded modulation. In section V we investigate the general cooperative broadcast scenario, and obtain an explicit rate expression by applying TS-EAF to the general multi-step conference. Finally, section VI presents concluding remarks.

II. Time-Sharing for the Single-Relay Case

A. Definitions

First, a word about notation: we denote discrete random variables with capital letters e.g. X, Y, and their realizations with lower case letters x, y. A random variable X takes values in a set $\mathcal{X}$. We use $||\mathcal{X}||$ to denote the cardinality of a finite discrete set $\mathcal{X}$, and $p_X(x)$ denotes the probability distribution function (p.d.f.) of X on $\mathcal{X}$. For brevity we may omit the subscript X when it is obvious from the context. We denote vectors with boldface letters, e.g. $\mathbf{x}$, $\mathbf{y}$; the i'th element of a vector $\mathbf{x}$ is denoted by x_i and we use $\mathbf{x}_i^j$ where $i < j$ to denote $(x_i, x_{i+1}, ..., x_{j-1}, x_j)$. We use $A_\epsilon^{*(n)}(X)$ to denote the set of ϵ-strongly typical sequences w.r.t. distribution $p_X(x)$ on $\mathcal{X}$, as defined in [23, ch. 5.1] and $A_\epsilon^{(n)}(X)$ to denote the ϵ-weakly typical set as defined in [24, ch. 3].

We also have the following definitions:

Definition 1: The *discrete relay channel* is defined by two discrete input alphabets $\mathcal{X}$ and $\mathcal{X}_1$, two discrete output alphabets $\mathcal{Y}$ and $\mathcal{Y}_1$ and a probability density function $p(y, y_1|x, x_1)$ giving the probability distribution on $\mathcal{Y} \times \mathcal{Y}_1$ for each $(x, x_1) \in \mathcal{X} \times \mathcal{X}_1$. The relay channel is called *memoryless* if the probability of a block of n transmissions is given by $p(\mathbf{y}, \mathbf{y}_1|\mathbf{x}, \mathbf{x}_1) = \prod_{i=1}^n p\left(y_i, y_{1,i}|x_i, x_{1,i}\right)$.
In this paper we consider only the memoryless relay channel.

Definition 2: A $(2^{nR}, n)$ *code* for the relay channel consists of a source message set $\mathcal{W} = \left\{1, 2, ..., 2^{nR}\right\}$, a mapping function f at the encoder,

$$f : \mathcal{W} \mapsto \mathcal{X}^n,$$

a set of n relay functions

$$x_{1,i} = t_i\left(y_{1,1}, y_{1,2}, ..., y_{1,i-1}\right),$$

where the i'th relay function t_i maps the first $i-1$ channel outputs at the relay into a transmitted relay symbol at time i. Lastly we have a decoder

$$g : \mathcal{Y}^n \mapsto \mathcal{W}.$$

Definition 3: The *average probability of error* for a code of length n for the relay channel is defined as

$$P_e^{(n)} = \Pr(g(Y^n) \neq W),$$

where W is selected uniformly over $\mathcal{W}$.

Definition 4: A rate R is called *achievable* if there exists a sequence of $(2^{nR}, n)$ codes with $P_e^{(n)} \rightarrow 0$ as $n \rightarrow \infty$.

B. The Single Relay EAF with Time-Sharing

Consider the following assignment of the auxiliary random variable of theorem 2:

$$p(\hat{y}_1|y_1, x_1) = \begin{cases} q & , \hat{y}_1 = y_1 \\ 1-q & , \hat{y}_1 = \Omega \notin \mathcal{Y}_1. \end{cases} \tag{5}$$

Under this assignment, the feasibility condition of (4) becomes

$$\begin{aligned}
I(X_1;Y) &\geq I(Y_1;\hat{Y}_1|X_1,Y)\\
&= H(Y_1|X_1,Y) - H(Y_1|X_1,Y,\hat{Y}_1)\\
&= H(Y_1|X_1,Y) - (1-q)H(Y_1|X_1,Y) - qH(Y_1|X_1,Y,Y_1)\\
&= qH(Y_1|X_1,Y),
\end{aligned}$$

and the rate expression (3) becomes

$$\begin{aligned}
R &\leq I(X;Y,\hat{Y}_1|X_1)\\
&= I(X;Y|X_1) + I(X;\hat{Y}_1|X_1,Y)\\
&= I(X;Y|X_1) + H(X|X_1,Y) - H(X|X_1,Y,\hat{Y}_1)\\
&= I(X;Y|X_1) + H(X|X_1,Y) - (1-q)H(X|X_1,Y) - qH(X|X_1,Y,Y_1)\\
&= I(X;Y|X_1) + qI(X;Y_1|X_1,Y).
\end{aligned}$$

Clearly, maximizing the rate implies maximizing q subject to the constraint $q \in [0,1]$. This gives the following corollary to theorem 2:

Corollary 1: For the general relay channel any rate R satisfying

$$R \leq I(X;Y|X_1) + \left[\frac{I(X_1;Y)}{H(Y_1|X_1,Y)}\right]^* I(X;Y_1|X_1,Y), \tag{6}$$

for the joint distribution $p(x,x_1,y,y_1) = p(x)p(x_1)p(y,y_1|x,x_1)$, with $[x]^* \triangleq \min(x,1)$, is achievable.

Now, consider the following distribution chain:

$$p(x,x_1,y,y_1,\hat{y}_1,\hat{\hat{y}}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)p(\hat{\hat{y}}_1|\hat{y}_1). \tag{7}$$

We note that this extended chain can be put into the standard form by letting $p(\hat{\hat{y}}_1|x_1,y_1) = \sum_{\hat{\mathcal{Y}}_1} p(\hat{y}_1,\hat{\hat{y}}_1|x_1,y_1) = \sum_{\hat{\mathcal{Y}}_1} p(\hat{y}_1|x_1,y_1)p(\hat{\hat{y}}_1|\hat{y}_1)$. After compression of Y_1 into $\hat{Y}_1$, there is a second compression operation, compressing $\hat{Y}_1$ into $\hat{\hat{Y}}_1$. The output of the second compression is used to facilitate cooperation between the relay and the destination. Therefore, the receiver decodes the message based on $\hat{\hat{\mathbf{y}}}_1$ and $\mathbf{y}$, repeating exactly the same step as in the standard relay decoding, with $\hat{\hat{\mathbf{y}}}$ replacing $\hat{\mathbf{y}}$. Then, the expressions of theorem 2 become

$$R \leq I(X;Y,\hat{\hat{Y}}_1|X_1), \tag{8}$$

$$\text{subject to } I(X_1;Y) \geq I(Y_1;\hat{\hat{Y}}_1|X_1,Y). \tag{9}$$

Now, applying TS to $\hat{\hat{Y}}_1$ with

$$p(\hat{\hat{y}}_1|\hat{y}_1) = \begin{cases} q & ,\hat{\hat{y}}_1 = \hat{y}_1 \\ 1-q & ,\hat{\hat{y}}_1 = \Delta \notin \hat{\mathcal{Y}}_1 \end{cases}, \tag{10}$$

the expressions in (8) and (9) become

$$
\begin{aligned}
R &\leq I(X;Y|X_1) + I(X;\hat{\hat{Y}}_1|X_1,Y) \\
&= I(X;Y|X_1) + H(X|X_1,Y) - H(X|\hat{\hat{Y}}_1,X_1,Y) \\
&= I(X;Y|X_1) + q(H(X|X_1,Y) - H(X|\hat{Y}_1,X_1,Y)) \\
&= I(X;Y|X_1) + qI(X;\hat{Y}_1|X_1,Y), \qquad (11) \\
I(X_1;Y) &\geq I(Y_1;\hat{\hat{Y}}_1|X_1,Y) \\
&= H(Y_1|X_1,Y) - H(Y_1|\hat{\hat{Y}}_1,X_1,Y) \\
&= H(Y_1|X_1,Y) - (1-q)H(Y_1|X_1,Y) - qH(Y_1|\hat{Y}_1,X_1,Y) \\
&= qI(Y_1;\hat{Y}_1|X_1,Y). \qquad (12)
\end{aligned}
$$

Combining this with the constraint $q \in [0,1]$ we obtain the following corollary to theorem 2:

Proposition 1: For the general relay channel, any rate R satisfying

$$R \leq I(X;Y|X_1) + \left[\frac{I(X_1;Y)}{I(Y_1;\hat{Y}_1|X_1,Y)}\right]^* I(X;\hat{Y}_1|X_1,Y),$$

for some joint distribution $p(x,x_1,y,y_1,\hat{y}_1) = p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$, is achievable.

This proposition generalizes on corollary 1 by performing a general Wyner-Ziv (WZ) compression combined with TS (which is a specific type of WZ compression), intended to guarantee feasibility of the first compression step. In section IV we apply a similar idea to the EAF relaying in the Gaussian relay channel scenario with coded modulation. Before we discuss the relationship between joint-decoding and time-sharing we present an intuitive way to view the EAF strategy.

C. An Intuitive View of Estimate-and-Forward

Consider the rate bound and the feasible region of theorem 2 given in equations (3) and (4). We note that the following intuitive explanation does not constitute a proof but it does provide an insight into the relay achievability results. We emphasize that the achievable rates stated in this section can also be proved rigorously. In the following we provide an intuitive insight into these expressions in terms of a flow on a graph.

In constructing the intuitive information flow representation for the relay channel, we first need to specify the underlaying assumptions and the operations performed at the source, the relay and the destination receiver:

- The source and the relay generate their codebooks independently.
- The relay compresses its channel output $\mathbf{y}_1$ into $\hat{\mathbf{y}}_1$, which represents the information conveyed to the destination receiver to assist in decoding the source message.
- Based on the above two restrictions we have the following Markov chain: $p(x)p(x_1)x(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$.
- The relay input signal $\mathbf{x}_1$ is based only on the compressed $\hat{\mathbf{y}}_1$.
- The destination uses $\mathbf{x}_1$, $\hat{\mathbf{y}}_1$ and $\mathbf{y}$ to decode the source message $\mathbf{x}$.

We also use the following representation for transmission, reception and compression:

- We represent an information source as a source whose output flow is equal to its information rate.
- We represent the compression operation as a flow sink whose flow consumption is equal to the mutual information between the original and the compressed sequences.
- The destination is represented as a flow sink.
- As in a standard flow on a graph, the flows are additive, following the chain rule of mutual information.

Now consider the following flow diagram of figure 2. As can be observed from the figure, the source has an

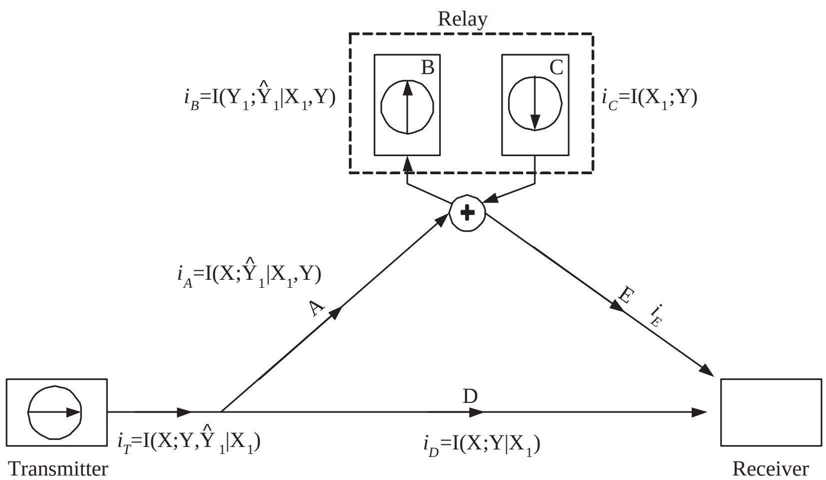

Fig. 2. The information flow budget for the general relay channel with compression at the relay.

output flow of

$$i_T = I(X; Y, \hat{Y}_1, X_1) = I(X; Y, \hat{Y}_1|X_1).$$

This follows from the fact that the destination uses $\mathbf{x}_1, \hat{\mathbf{y}}_1$ and $\mathbf{y}$ to decode $\mathbf{x}$ and the fact that X and X_1 are independent. This total flow reaches the receiver through two branches, the direct branch (D) which carries a flow of $i_D = I(X; Y|X_1)$ and the relay branch (ABCE). Now, the quantities in the relay branch are calculated given X_1 and Y to represent only the rate increase over the direct path. The relay branch has four parts: an edge (A) which carries a flow of $I(X; \hat{Y}_1|X_1, Y)$, a sink (B) with consumption $I(Y_1; \hat{Y}_1|X_1, Y)$, a relay source (C) with an output flow of $I(X_1; Y)$ and an edge (E) from the relay to the destination. Here, the relay transmission to the destination (C) is done at a fixed rate $I(X_1; Y)$, independent of the type of compression $p(\hat{y}_1|x_1)$ used at the relay, since we always transmit from the relay to the destination at the maximum possible rate in order to obtain the best performance. The rate loss due to compression is represented by $I(\hat{Y}_1; Y_1|X_1, Y)$, since we consider only the excess rates over the direct one.

Now, from the laws of flow addition and conservation, the overall flow from the source to the destination through the relay branch is $i_E = i_A + i_B + i_C$. To assist the direct link (D) we need the flow on (ABCE) to be positive.

In theorem 2 the scheme considers only the last two elements, $i_B + i_C$, and verifies that their net flow is positive, namely

$$-I(Y_1; \hat{Y}_1 | X_1, Y) + I(X_1; Y) > 0. \tag{13}$$

This condition guarantees a net positive flow on (ABCE) since always $i_A \geq 0$. Now, the flow to the destination can be obtained as the minimum

$$R \leq \min \{ i_D + i_E, i_T \}, \tag{14}$$

where, the second term in the minimum is obtained from the transmitter, since trivially the information rate at the receiver cannot exceed i_T. We note that because $i_B + i_C \geq 0$, the minimum in (14) is i_T. Therefore, the resulting achievable rate is

$$R \leq I(X; Y, \hat{Y}_1 | X_1),$$

which combined with (13) gives the result of [2, theorem 6].

However, the condition in (13) is not tight since even when $i_B + i_C < 0$ the flow on (ABCE) is still non-negative if the entire sum $i_A + i_B + i_C$ is non-negative, i.e.

$$I(X; \hat{Y}_1 | X_1, Y) - I(\hat{Y}_1; Y_1 | X_1, Y) + I(X_1; Y) \geq 0. \tag{15}$$

Then, the achievable rate to the destination is bounded by

$$R \leq i_D + i_E = I(X; Y | X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y). \tag{16}$$

Indeed, when the flow through the relay branch (ABCE) is zero we obtain the non-cooperative rate $I(X; Y | X_1)$. Plugging the expression (16) into (14) yields the following achievable rate:

$$\begin{aligned} R &\leq \min \{ i_D + i_E, i_T \} \\ &= \min \left\{ I(X; Y | X_1) + I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y), I(X; Y, \hat{Y}_1 | X_1) \right\} \\ &= I(X; Y | X_1) + \min \left\{ I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y), I(X; \hat{Y}_1 | X_1, Y) \right\}. \end{aligned}$$

Combining this with (15), (informally) proves the following proposition:

Proposition 2: For the general relay channel, any rate R *satisfying*

$$\begin{aligned} R &\leq I(X; Y | X_1) + \min \left\{ I(X_1; Y) - I(\hat{Y}_1; Y_1 | X, X_1, Y), I(X; \hat{Y}_1 | X_1, Y) \right\}, \\ \textit{subject to } I(X_1; Y) &\geq I(\hat{Y}_1; Y_1 | X, X_1, Y) = I(\hat{Y}_1; Y_1 | X_1, Y) - I(X; \hat{Y}_1 | X_1, Y), \end{aligned}$$

for some joint distribution $p(x, x_1, y, y_1, \hat{y}_1) = p(x)p(x_1)p(y, y_1 | x, x_1)p(\hat{y}_1 | x_1, y_1)$*, is achievable.*

The proof of proposition 2 can be made formal using joint-decoding at the destination receiver, but as in the next subsection we show that this expression is a special case of [2, theorem 6] obtained by time-sharing, we omit the details of the proof here.

D. Joint-Decoding and Time-Sharing

In the original work of [2, theorem 6], the decoding procedure at the destination receiver for decoding the message w_{i-1} at time i is composed of three steps (the notations below are identical to [2, theorem 6]. The reader is referred to the proof of [2, theorem 6] to recall the definitions of the sets and variables used in the following description):

1) Decode the relay index s_i using $\mathbf{y}(i)$, the received signal at time i.
2) Decode the relay message z_{i-1}, using s_i, the received signal $\mathbf{y}(i-1)$ and the previously decoded s_{i-1}.
3) Decode the source message w_{i-1} using $\mathbf{y}(i-1)$, z_{i-1} and s_{i-1}.

Evidently, when decoding the relay message z_{i-1} at the second step, the receiver does not make use of the statistical dependence between $\hat{\mathbf{y}}_1(i-1)$, the relay sequence at time $i-1$, and $\mathbf{x}(w_{i-1})$, the transmitted source codeword at time $i-1$. The way to use this dependence is to jointly decode z_{i-1} and w_{i-1} after decoding s_i and s_{i-1}. The joint-decoding procedure then has the following steps:

1) From $\mathbf{y}(i)$, the received signal at time i, the receiver decodes s_i by looking for a unique $s \in \mathcal{S}$, the set of indices used to select $\mathbf{x}_1$, such that $\big(\mathbf{x}_1(s), \mathbf{y}(i)\big) \in A_\epsilon^{*(n)}$. As in [2, theorem 6], the correct s_i can be decoded with an arbitrarily small probability of error by taking n large enough as long as
$$R_0 \leq I(X_1;Y), \tag{17}$$
where $||\mathcal{S}|| = 2^{nR_0}$.
2) The receiver now knows the set S_{s_i} into which z_{i-1} (the relay message at time $i-1$) belongs. Additionally, from decoding at time $i-1$ the receiver knows s_{i-1}, used to generate z_{i-1}.
3) The receiver generates the set $\mathcal{L}(i-1) = \Big\{w \in \mathcal{W} : \big(\mathbf{x}(w), \mathbf{y}(i-1), \mathbf{x}_1(s_{i-1})\big) \in A_\epsilon^{*(n)}\Big\}$.
4) The receiver now looks for a unique $w \in \mathcal{L}(i-1)$ such that $\big(\mathbf{x}(w), \mathbf{y}(i-1), \hat{\mathbf{y}}_1(z|s_{i-1}), \mathbf{x}_1(s_{i-1})\big) \in A_\epsilon^{*(n)}$ for some $z \in S_{s_i}$. If such a unique w exists then it is the decoded $\hat{w}_{i-1}$, otherwise the receiver declares an error.

We do not give here a formal proof for the resulting rate expression, but as indicated in section II-C, the rate expression resulting from this decoding procedure is given by proposition 2.

Let us now compare the the rates obtained with joint-decoding (proposition 2) with the rates obtained with the sequential decoding of [2, thoerem 6]: to that end we consider the joint-decoding result of proposition 2 with the extended probability chain of (7):

$$p(x, x_1, y, y_1, \hat{y}_1, \hat{\hat{y}}_1) = p(x)p(x_1)p(y, y_1|x, x_1)p(\hat{y}_1|x_1, y_1)p(\hat{\hat{y}}_1|\hat{y}_1),$$

where $\hat{\hat{Y}}_1$ represents the information relayed to the destination. Expanding the expressions of proposition 2 using the assignment (10), similarly to proposition 1, we obtain the expressions:

$$R \leq I(X;Y|X_1) + \min\Big\{I(X_1;Y) - qI(\hat{Y}_1;Y_1|X, X_1, Y), qI(X;\hat{Y}_1|X_1, Y)\Big\} \tag{18}$$
$$\text{subject to } I(X_1;Y) \geq qI(\hat{Y}_1;Y_1|X, X_1, Y) = q\Big(I(\hat{Y}_1;Y_1|X_1, Y) - I(X;\hat{Y}_1|X_1, Y)\Big). \tag{19}$$

We can now make the following observations:

1) Setting $q = 1$ we obtain proposition 2. Additionally, if $I(X_1;Y) > I(\hat{Y}_1;Y_1|X_1,Y)$ then both proposition 2 and [2, theorem 6] give identical expressions.
2) When $q = 1$ and
$$I(\hat{Y}_1;Y_1|X_1,Y) - I(X;\hat{Y}_1|X_1,Y) < I(X_1;Y) < I(\hat{Y}_1;Y_1|X_1,Y), \tag{20}$$
then *for the same* mapping $p(\hat{y}_1|x_1,y_1)$ we obtain that proposition 2 provides rate but [2, theorem 6] does not. The rate expression under these conditions is
$$R \leq I(X;Y|X_1) + I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y). \tag{21}$$
3) Now, fix the probability chain $p(x)p(x_1)p(y,y_1|x,x_1)p(\hat{y}_1|x_1,y_1)$ and examine the expressions (18) and (19) when (20) holds: when $q < 1$, then (20) guarantees that condition (19) is still satisfied. If q is close enough to 1 such that we also have $I(X_1;Y) \leq qI(\hat{Y}_1;Y_1|X_1,Y)$, the rate from (18), i.e.,
$$R \leq I(X;Y|X_1) + I(X_1;Y) - qI(\hat{Y}_1;Y_1|X,X_1,Y),$$
is now greater than (21). In this case can keep decreasing q until
$$I(X_1;Y) - qI(\hat{Y}_1;Y_1|X,X_1,Y) = qI(X;\hat{Y}_1|X_1,Y) \tag{22}$$
at which point the rate becomes
$$R \leq I(X;Y|X_1) + qI(X;\hat{Y}_1|X_1,Y). \tag{23}$$
This rate can be obtained from [2, theorem 6] by applying the extended probability chain of (7), as long as $I(X_1;Y) \geq qI(\hat{Y}_1,Y_1|X_1,Y)$.

We therefore conclude that all the rates that joint decoding allows can also be obtained or exceeded by the original EAF with an appropriate time sharing[1].

Note that equality in (22) implies
$$q_{opt} = \min\left\{1, \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X,X_1,Y) + I(X;\hat{Y}_1|X_1,Y)}\right\} = \min\left\{1, \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}\right\},$$
hence q_{opt} is the maximum q that makes the mapping $p(\hat{y}_1|x_1,y_1)$ feasible for [2, theorem 6]. Plugging q_{opt} into (23), we obtain the rate expression of proposition 1.

Finally, consider again the region where joint decoding is useful (20):
$$\begin{aligned}
I(\hat{Y}_1;Y_1|X,X_1,Y) &\leq I(X_1;Y) \leq I(\hat{Y}_1;Y_1|X_1,Y)\\
\Rightarrow 0 &\leq I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y) \leq I(\hat{Y}_1;Y_1|X_1,Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)\\
\Rightarrow 0 &\leq I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y) \leq I(X_1;\hat{Y}_1|X_1,Y)\\
\Rightarrow 0 &\leq \tfrac{I(X_1;Y) - I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \leq 1.
\end{aligned}$$

[1]This argument is due to Shlomo Shamai and Gerhard Kramer.

If $I(X;\hat{Y}_1|X_1,Y)>0$, then using time-sharing on $\hat{Y}_1$ with

$$q=\frac{I(X_1;Y)-I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} \tag{24}$$

into equations (11) and (12) yields:

$$I(X;Y|X_1)+qI(X;\hat{Y}_1|X_1,Y)=I(X;Y|X_1)+I(X_1;Y)-I(\hat{Y}_1;Y_1|X,X_1,Y),$$

as long as $I(X_1;Y)\geq qI(\hat{Y}_1;Y_1|X_1,Y)$, or equivalently

$$q\leq\frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}. \tag{25}$$

Plugging assignment (24) into (25) we obtain:

$$\begin{aligned}
\frac{I(X_1;Y)-I(\hat{Y}_1;Y_1|X,X_1,Y)}{I(X;\hat{Y}_1|X_1,Y)} &\leq \frac{I(X_1;Y)}{I(\hat{Y}_1;Y_1|X_1,Y)}\\
\Rightarrow\quad \Big(I(X_1;Y)-I(\hat{Y}_1;Y_1|X,X_1,Y)\Big)I(\hat{Y}_1;Y_1|X_1,Y) &\leq I(X_1;Y)I(X;\hat{Y}_1|X_1,Y)\\
\Rightarrow\quad I(X_1;Y)I(\hat{Y}_1;Y_1|X_1,Y)-I(X_1;Y)I(X;\hat{Y}_1|X_1,Y) &\leq I(\hat{Y}_1;Y_1|X,X_1,Y)I(\hat{Y}_1;Y_1|X_1,Y)\\
\Rightarrow\quad I(X_1;Y)I(\hat{Y}_1;Y_1|X,X_1,Y) &\leq I(\hat{Y}_1;Y_1|X,X_1,Y)I(\hat{Y}_1;Y_1|X_1,Y)\\
\Rightarrow\quad I(X_1;Y) &\leq I(\hat{Y}_1;Y_1|X_1,Y),
\end{aligned}$$

as long as $I(\hat{Y}_1;Y_1|X,X_1,Y)>0$, which is the region where joint-decoding is supposed to be useful. Hence the joint-decoding rate of proposition 2 can be obtained by time sharing on the [2, theorem 6] expression. Therefore, joint-decoding does not improve on the rate of [2, theorem 6]. In fact the rate of proposition 1 is always at least as large as that of proposition 2.

III. An Achievable Rate for the Relay Channel with Multiple Relays

When the source-relay channel is very noisy then, as discussed in the introduction, it may be better not to use the relay at all than to employ the decode-and-forward strategy. Alternatively, when decode-and-forward is not useful, one could employ estimate-and-forward. One result for multiple relays based on EAF can be found in [9] which considered the two-relay case. In [8, theorem 3] the EAF strategy, with partial decoding was applied to the multiple-relay case, and in [8, theorem 4] a mixed EAF and DAF strategy was applied. However, as stated in [8, remark 22, remark 23] applying the general estimate-and-forward to a network with an arbitrary number of relays is computationally impractical due to the large number of constraints that characterize the feasible region (for two relays we need to satisfy 9 constraints). Moreover, the rate computation is prohibitive since it would imply solving a non-convex optimization problem. In conclusion, an alternative achievable rate to that based on decode-and-forward, which can also be evaluated with a reasonable effort, has not been presented to date. In this section we derive an explicit achievable rate based on estimate-and-forward. The strategy we use is to pick the auxiliary random variable such that the feasibility constraints are satisfied. This is not a trivial choice since setting the auxiliary random variable in theorem 2 to be the relay channel output (i.e. $\hat{Y}_1=Y_1$) does not remove this constraint, and we therefore need to incorporate time-sharing as discussed in the following.

A. A General Achievable Rate

We extend the idea of section II-B to the relay channel with N relays. This channel consists of a source with channel input X, N relays where for relay i, X_i denotes the channel input and Y_i denotes the channel output, and a destination with channel output Y. This channel is denoted by $\left(\mathcal{X} \times_{i=1}^{N} \mathcal{X}_i, p(y, y_1, ..., y_N|x, x_1, ..., x_N), \mathcal{Y} \times_{i=1}^{N} \mathcal{Y}_i\right)$. Let $\mathbf{X} = (X_1, X_2, ..., X_N)$ and $\mathbf{Y} = (Y_1, Y_2, ..., Y_N)$. We now have the following theorem:

Theorem 3: For the general multiple-relay channel with N relays, $\Big(\mathcal{X} \times_{i=1}^{N} \mathcal{X}_i, p(y, y_1, ..., y_N|x, x_1, ..., x_N),$ $\mathcal{Y} \times_{i=1}^{N} \mathcal{Y}_i\Big)$, any rate R satisfying

$$R \leq I(X;Y|\mathbf{X}) + \sum_{\theta=1}^{2^N-1} P(\mathit{Bin}_N(\theta)) I(X; \mathbf{Y}_{\mathit{Bin}_N(\theta)}|\mathbf{X}, Y),$$

where $\mathit{Bin}_N(\theta)$ is an N-element vector that contains $'1'$ in the locations where the N-bit binary representation of the integer θ contains $'1'$, $P(\mathit{Bin}_N(\theta)) = \prod_{i:\mathit{Bin}_N(\theta)_i=0}(1-q_i)\prod_{i:\mathit{Bin}_N(\theta)_i=1} q_i$, $\mathit{Bin}_N(\theta)_i$ is the i'th bit in the N-bit binary representation of θ, $\mathbf{Y}_{\mathit{Bin}_N(\theta)} = (Y_{i_1}, Y_{i_2}, ..., Y_{i_M})$, where i_1, i_2, ..., i_M are the locations of the $'1'$ in $\mathit{Bin}_N(\theta)$, and

$$q_i = \left[\frac{I(X_i;Y|\tilde{\mathbf{Z}}_i)}{H(Y_i|\mathbf{X}, Y) - \sum_{j=1}^{2^{L'_i}-1} P_{l'}(\mathit{Bin}_{L'_i}(j)) I(Y_i; \tilde{\mathbf{Y}}_{l', \mathit{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i)|\mathbf{X}, Y)}\right]^*, \tag{26}$$

for the joint distribution $p(x, x_1, x_2, ..., x_N, y, y_1, y_2, ..., y_N) = p(x)p(x_1)...p(x_N)p(y, y_1, ..., y_N|x, x_1, ..., x_N)$ is achievable. In (26) *$\tilde{\mathbf{Z}}_i$ is the vector containing all the variables X_j decoded prior to decoding X_i, $\tilde{\mathbf{T}}_i$ is a vector that contains all the variables $\hat{Y}_p$ decoded prior to decoding $\hat{Y}_i$, and $\tilde{\mathbf{Y}}_{l', \mathit{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i)$ contains all the $Y_{l'_r}$, such that $\hat{Y}_{l'_r} \in \tilde{\mathbf{T}}_i$, and r is a location of $'1'$ in the L'_i-bit binary representation of j. L'_i if the number of elements in $\tilde{\mathbf{T}}_i$. Note that if $\hat{Y}_p \in \tilde{\mathbf{T}}_i$ then we must have $X_p \in \tilde{\mathbf{Z}}_i$.*

To facilitate the understanding of the expressions in theorem 3, we first look at a simplified case where the destination decodes each relay message independently of the messages of the other relays. This can be obtained from theorem 3 by setting $\tilde{\mathbf{Z}}_i = \varnothing$ and $\tilde{\mathbf{T}}_i = \varnothing$, $i = 1, 2, ..., N$. The result is summarized in the following corollary:

Corollary 2: For the general multiple-relay channel $\left(\mathcal{X} \times_{i=1}^{N} \mathcal{X}_i, p(y, y_1, ..., y_N|x, x_1, ..., x_N), \mathcal{Y} \times_{i=1}^{N} \mathcal{Y}_i\right)$, any rate R satisfying

$$R \leq I(X;Y|\mathbf{X}) + \sum_{\theta=1}^{2^N-1} P(\mathit{Bin}_N(\theta)) I(X; \mathbf{Y}_{\mathit{Bin}_N(\theta)}|\mathbf{X}, Y), \tag{27}$$

is achievable, where

$$q_i = \left[\frac{I(X_i;Y)}{H(Y_i|\mathbf{X}, Y)}\right]^*, \tag{28}$$

for the joint distribution $p(x, x_1, x_2, ..., x_N, y, y_1, y_2, ..., y_N) = p(x)p(x_1)...p(x_N)p(y, y_1, ..., y_N|x, x_1, ..., x_N)$.

In the multi-relay strategy we employ in this section each relay transmits its channel output Y_i with probability q_i, independent of the other relays. Therefore, when considering a group of N relays, the probability that any subgroup of relays will transmit their channel outputs simultaneously is simply the product of all transmission

probabilities q_i at each relay in the group, multiplied by the product of erasure probabilities $(1-q_i)$ for each relay in the complement group. Now, considering the rate expression of (27) we observe that the rate is obtained by taking all possible groupings of relays. For each grouping the resulting rate is the rate obtained when using all the channel outputs of all the relays in that group to assist in decoding. This is indicated by the term $\mathbf{Y}_{\mathrm{Bin}_N(\theta)}$. This rate has to be weighted by the probability of such an overlap occurring, which is given by $P(\mathrm{Bin}_N(\theta))$. We then sum over all such groupings to obtain the achievable rate. The parameter q_i for each relay, which is determined by (28), can be interpreted by considering the terms in the denominator and numerator: the denominator $H(Y_i|\mathbf{X},Y)$ is the (exponent of the) size of uncertainty at the destination receiver about relay i's output Y_i. The numerator is the (exponent of the) size of the information set that can be transmitted from relay i to the destination receiver. Therefore, the fraction $\frac{I(X_i;Y)}{H(Y_i|\mathbf{X},Y)}$ can be interpreted as the maximal fraction of the uncertainty at the destination about relay i's channel output Y_i, that can be compensated by the relay transmission. Of course, this faction has to be upper bounded by one. In the more general setup of theorem 3, the decoding of the relay information from relay i is done by using the information from the relays which were decoded before relay i to assist in decoding. This results in the conditioning at the numerator and the negative terms in the denominator, both contribute to increasing the value of q_i.

B. *Proof of Theorem 3*

1) Overview of Coding Strategy: The transmitter generates its codebook independent of the relays. Next, each relay generates its own codebook independent of the other relays following the construction of [2, theorem 6], with the mapping $p(\hat{y}_i|x_i,y_i)$ at each relay set to the time-sharing mapping of (5) with parameter q_i. The destination receiver first needs to decode all the relay codewords $\{X_i^n\}_{i=1}^N$ and use this information to decode the relay messages $\left\{\hat{Y}_i^n\right\}_{i=1}^N$. To this end, the relay decides on a decoding order for the X_i^n sequences and a decoding order for the $\hat{Y}_i^n$ sequences. These decoding orders determine the maximum value of q_i that can be selected for each relay, thereby allowing us to determine the auxiliary variables' mappings and obtain an explicit rate expression. Finally, the receiver uses all the decoded $\{X_i^n\}_{i=1}^N$ and $\left\{\hat{Y}_i^n\right\}_{i=1}^N$ sequences, together with its channel input to decode the source message.

We now give the details of the construction: fix the distributions $p(x)$, $p(x_1)$, $p(x_2)$,...,$p(x_N)$, and

$$p(\hat{y}_i|x_i,y_i)=\begin{cases} q_i & ,\hat{y}_i=y_i \\ 1-q_i & ,\hat{y}_i=\Omega\notin\mathcal{Y}_i \end{cases}, \tag{29}$$

$i=1,2,...,N$. Let $\mathcal{W}=\left\{1,2,...,2^{nR}\right\}$ be the source message set.

2) Code Construction at the Transmitter and the Relays:

- Code construction and transmission at the transmitter are the same as in [2, theorem 6].
- Code construction at the relays is done by repeating the relay code construction of [2, theorem 6] for each relay, where relay i uses the distributions $p(\hat{y}_i|x_i,y_i)$ and $p(x_i)$. We denote the relay message, the transmitted message and the partition set at relay i at time k with $z_{i,k}$, $s_{i,k}$ and $S^{(i)}_{s_{i,k}}$ respectively. The message set for s_i is

denoted $\mathcal{W}_i$, where $||\mathcal{W}_i|| = 2^{nR_i}$. The message set for z_i is denoted $\mathcal{W}'_i$, $||\mathcal{W}'_i|| = 2^{nR'_i}$. The relay codewords at relay i are denoted $\hat{\mathbf{y}}_i(z_i|s_i)$, and the transmitted codewords at relay i are denoted $\mathbf{x}_i(s_i)$, $s_i \in \mathcal{W}_i$, $z_i \in \mathcal{W}'_i$.

3) Decoding and Encoding at the Relays:

Consider relay i at time $k-1$:

- From the relay transmission at time $k-1$, the relay knows $s_{i,k-1}$. Now the relay looks for a message $z_i \in \mathcal{W}'_i$, such that
$$\left(\hat{\mathbf{y}}_i(z_i|s_{i,k-1}), \mathbf{y}_i(k-1), \mathbf{x}_i(s_{i,k-1})\right) \in A_\epsilon^{*(n)}(\hat{Y}_i, Y_i, X_i).$$
Following the argument in [2, theorem 6], for n large enough there is such a message z_i with a probability that is arbitrarily close to 1, as long as
$$R'_i > I(\hat{Y}_i; Y_i|X_i) + \epsilon = q_i H(Y_i|X_i) + \epsilon. \tag{30}$$
Denote this message with $z_{i,k-1}$.
- Let $s_{i,k}$ be the index of the partition of $\mathcal{W}'_i$ into which $z_{i,k-1}$ belongs, i.e., $z_{i,k-1} \in S^{(i)}_{s_{i,k}}$.
- At time k relay i transmits $\mathbf{x}_i(s_{i,k})$.

4) Decoding at the Destination:

- Consider the decoding of w_{k-1} at time k, for a fixed decoding order: let $\tilde{\mathbf{Z}}_i$ contain all the X_j's whose $s_{j,k}$'s are decoded prior to decoding $s_{i,k}$. Therefore, decoding $s_{i,k}$ is done by looking for a unique message $s_i \in \mathcal{W}_i$ such that
$$\left(\mathbf{x}_i(s_i), \mathbf{x}_{m_1}(s_{m_1,k}), \mathbf{x}_{m_2}(s_{m_2,k}), ..., \mathbf{x}_{m_{M_i}}(s_{m_{M_i},k}), \mathbf{y}(k)\right) \in A_\epsilon^{*(n)}(X_i, \tilde{\mathbf{Z}}_i, Y),$$
where m_1, m_2,...,m_{M_i} enumerate all the X_j's in $\tilde{\mathbf{Z}}_i = \left(X_{m_1}, X_{m_2}, ...X_{m_{M_i}}\right)$. Assuming correct decoding at the previous steps, then by the point-to-point channel achievability proof we obtain that the probability of error for decoding $s_{i,k}$ can be made arbitrarily small by taking n large enough as long as
$$R_i < I(X_i; Y, \tilde{\mathbf{Z}}_i) - \epsilon = I(X_i; Y|\tilde{\mathbf{Z}}_i) - \epsilon. \tag{31}$$
Let $\tilde{\mathbf{T}}_i$ contain all the $\hat{Y}_{l'}$'s whose $z_{l',k-1}$'s are decoded prior to decoding $z_{i,k-1}$. Note that all the $\{s_{i,k-1}\}_{i=1}^N$ were already decoded at the previous time interval when w_{k-2} was decoded.
- The destination generates the set
$$\begin{aligned}\mathcal{L}_i(k-1) = \Big\{ z_i \in \mathcal{W}'_i : \big(&\mathbf{y}(k-1), \hat{\mathbf{y}}_i(z_i|s_{i,k-1}), \hat{\mathbf{y}}_{l'_1}(z_{l'_1,k-1}|s_{l'_1,k-1}), ..., \hat{\mathbf{y}}_{l'_{L'_i}}(z_{l'_{L'_i},k-1}|s_{l'_{L'_i},k-1}),\\ &\mathbf{x}_1(s_{1,k-1}), \mathbf{x}_2(s_{2,k-1}), ..., \mathbf{x}_N(s_{N,k-1})\big) \in A_\epsilon^{*(n)}(Y, \hat{Y}_i, \tilde{\mathbf{T}}_i, \mathbf{X}) \Big\},\end{aligned} \tag{32}$$
where l'_1, l'_2,...,l'_{L_i} enumerate all the $\hat{Y}_{l'}$'s in $\tilde{\mathbf{T}}_i$. The average size of $\mathcal{L}_i(k-1)$ can be bounded using the standard technique of [2, equation (36)] and the fact that when $z_i \neq z_{i,k-1}$, then the corresponding $\hat{\mathbf{y}}_i(z_i|s_{i,k-1})$ is independent of all the variables in (32) except $\mathbf{x}_i(s_{i,k-1})$. The resulting bound is
$$E\left\{||\mathcal{L}_i(k-1)||\right\} \leq 1 + 2^{n(R'_i - I(\hat{Y}_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i|X_i) + 3\epsilon)},$$

where $\mathbf{X}_{-i}$ is an $N-1$ element vector that contains all the elements of $\mathbf{X}$ except X_i.

- Now, the destination looks for a unique $z_i \in \mathcal{L}_i(k-1) \bigcap S^{(i)}_{s_{i,k}}$. Therefore, making the probability of error arbitrarily small by taking n large enough can be done as long as

$$R'_i < I(\hat{Y}_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) + I(X_i; Y | \tilde{\mathbf{Z}}_i) - 4\epsilon. \tag{33}$$

We note that using the assignment (29) we can write

$$\begin{aligned}
I(\hat{Y}_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) &= H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) - H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i, \hat{Y}_i) \\
&= H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) - (1-q_i) H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) - q_i H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i, Y_i) \\
&= q_i H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) - q_i H(Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i, Y_i) \\
&= q_i I(Y_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}_i | X_i) \\
&= q_i \left(H(Y_i | X_i) - H(Y_i | Y, \mathbf{X}_{-i}, X_i, \hat{Y}_{l'_1}, \tilde{\mathbf{T}}^{L'_i}_{i,2}) \right) \\
&= q_i \Big(q_{l'_1} H(Y_i | X_i) + (1 - q_{l'_1}) H(Y_i | X_i) \\
&\qquad - q_{l'_1} H(Y_i | Y, \mathbf{X}_{-i}, X_i, Y_{l'_1}, \tilde{\mathbf{T}}^{L'_i}_{i,2}) - (1 - q_{l'_1}) H(Y_i | Y, \mathbf{X}_{-i}, X_i, \tilde{\mathbf{T}}^{L'_i}_{i,2}) \Big) \\
&= q_i \Big(q_{l'_1} I(Y_i; Y, \mathbf{X}_{-i}, Y_{l'_1}, \tilde{\mathbf{T}}^{L'_i}_{i,2} | X_i) + (1 - q_{l'_1}) I(Y_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{T}}^{L'_i}_{i,2} | X_i) \Big) \\
&\dots \\
&= q_i \sum_{j=0}^{2^{L'_i}-1} P_{l'}(\text{Bin}_{L'_i}(j)) I(Y_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{Y}}_{l', \text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i) | X_i),
\end{aligned}$$

where $P_{l'}(\text{Bin}_{L'_i}(j)) = \prod_{r:\text{Bin}_{L'_i}(j)_r = 1} q_{l'_r} \times \prod_{r:\text{Bin}_{L'_i}(j)_r = 0} (1 - q_{l'_r})$, $\text{Bin}_{L'_i}(j)_r$ is the r-th bit of the L'_i-bit binary representation of j, and $\tilde{\mathbf{Y}}_{l', \text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i) = \left(Y_{l'_{n_1}}, Y_{l'_{n_2}}, ..., Y_{l'_{n_M}} \right)$, $n_1, n_2, ..., n_M$ are the locations of '1' in the L'_i-bit binary representation of j, and $l'_{n_1}, l'_{n_2}, ..., l'_{n_M}$ are the indices of the $\hat{Y}_i$'s in locations $n_1, n_2, ..., n_M$ in $\tilde{\mathbf{T}}_i$. For example, if $L'_i = 3$ and $j = 3$ then $\text{Bin}_3(3) = (1, 0, 1)$ and $M = 2$, $n_1 = 1, n_2 = 3$. Letting $\tilde{\mathbf{T}}_i = \left(\hat{Y}_3, \hat{Y}_1, \hat{Y}_2 \right)$ then $l'_1 = 3, l'_2 = 1$ and $l'_3 = 2$, and

$$\begin{aligned}
P_{l'}(\text{Bin}_3(3)) &= q_{l'_1}(1 - q_{l'_2}) q_{l'_3}, \\
\tilde{\mathbf{Y}}_{l', \text{Bin}_3(3)}(\tilde{\mathbf{T}}_i)) &= (Y_{l'_1}, Y_{l'_3}) = (Y_3, Y_2).
\end{aligned}$$

5) Combining the Bounds on R'_i: Applying the above scheme requires that R'_i satisfies (30) and (33):

$$q_i H(Y_i | X_i) + \epsilon < R'_i < q_i \sum_{j=0}^{2^{L'_i}-1} P_{l'}(\text{Bin}_{L'_i}(j)) I(Y_i; Y, \mathbf{X}_{-i}, \tilde{\mathbf{Y}}_{l', \text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i) | X_i) + I(X_i; Y | \tilde{\mathbf{Z}}_i) - 4\epsilon,$$

which is satisfied if

$$\begin{aligned}
q_i &< \frac{I(X_i;Y|\tilde{\mathbf{Z}}_i)-5\epsilon}{H(Y_i|X_i)-\sum_{j=0}^{2^{L'_i}-1} P_{l'}(\text{Bin}_{L'_i}(j))I(Y_i;Y,\mathbf{X}_{-i},\tilde{\mathbf{Y}}_{l',\text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i)|X_i)} \\
&= \frac{I(X_i;Y|\tilde{\mathbf{Z}}_i)-5\epsilon}{H(Y_i|X_i)-I(Y_i;Y,\mathbf{X}_{-i}|X_i)-\sum_{j=1}^{2^{L'_i}-1} P_{l'}(\text{Bin}_{L'_i}(j))I(Y_i;\tilde{\mathbf{Y}}_{l',\text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i)|\mathbf{X},Y)} \\
&= \frac{I(X_i;Y|\tilde{\mathbf{Z}}_i)-5\epsilon}{H(Y_i|\mathbf{X},Y)-\sum_{j=1}^{2^{L'_i}-1} P_{l'}(\text{Bin}_{L'_i}(j))I(Y_i;\tilde{\mathbf{Y}}_{l',\text{Bin}_{L'_i}(j)}(\tilde{\mathbf{T}}_i)|\mathbf{X},Y)}.
\end{aligned}$$

Combining with the constraint $0 \le q_i \le 1$ gives the condition in (26).

Finally, the achievable rate is obtained as follows: using the decoded $\{\hat{\mathbf{y}}_i(z_{i,k-1}|s_{i,k-1})\}_{i=1}^N$ (assuming correct decoding of all $\{z_{i,k-1}\}_{i=1}^N$) the receiver decodes the source message w_{k-1} by looking for a message $w \in \mathcal{W}$ such that

$$\Big(\mathbf{x}(w),\hat{\mathbf{y}}_1(z_{1,k-1}|s_{1,k-1}),\hat{\mathbf{y}}_2(z_{2,k-1}|s_{2,k-1}),...,,\hat{\mathbf{y}}_N(z_{N,k-1}|s_{N,k-1}),$$
$$\mathbf{x}_1(s_{1,k-1}),\mathbf{x}_2(s_{2,k-1}),...,\mathbf{x}_N(s_{N,k-1}),\mathbf{y}(k-1)\Big) \in A_\epsilon^{*(n)}(X,\hat{\mathbf{Y}},\mathbf{X},Y),$$

where $\hat{\mathbf{Y}} = \Big(\hat{Y}_1,\hat{Y}_2,...,\hat{Y}_N\Big)$. This results in an achievable rate of

$$R \le I(X;Y,\hat{\mathbf{Y}},\mathbf{X}) = I(X;Y,\hat{\mathbf{Y}}|\mathbf{X}).$$

Plugging in the assignments of all the $\hat{Y}_i$'s, we get the following explicit rate expression:

$$\begin{aligned}
I(X;Y,\hat{\mathbf{Y}}|\mathbf{X}) &= I(X;Y|\mathbf{X})+I(X;\hat{\mathbf{Y}}|\mathbf{X},Y) \\
&= I(X;Y|\mathbf{X})+H(X|\mathbf{X},Y)-H(X|\mathbf{X},Y,\hat{\mathbf{Y}}) \\
&= I(X;Y|\mathbf{X})+H(X|\mathbf{X},Y)-(1-q_1)H(X|\mathbf{X},Y,\hat{\mathbf{Y}}_2^N)-q_1H(X|\mathbf{X},Y,\hat{\mathbf{Y}}_2^N,Y_1) \\
&= I(X;Y|\mathbf{X})+(1-q_1)I(X;\hat{\mathbf{Y}}_2^N|\mathbf{X},Y)+q_1I(X;\hat{\mathbf{Y}}_2^N,Y_1|\mathbf{X},Y) \\
&\ldots \\
&= I(X;Y|\mathbf{X})+\sum_{\theta=1}^{2^N-1} P(\text{Bin}_N(\theta))I(X;\mathbf{Y}_{\text{Bin}_N(\theta)}|\mathbf{X},Y).
\end{aligned}$$

■

C. Discussion

To demonstrate the usefulness of the explicit EAF-based achievable rate of theorem 3 we compare it with the DAF-based method of [5, theorem 3.1] for the two-relay case. For this scenario there are five possible DAF setups,

and the maximum of the five resulting rates is taken as the DAF-based rate:

$$\begin{aligned}
R^{DAF} &= \sup_{p(x,x_1,x_2)} \max\{R_1, R_2, R_{12}, R_{21}, R_G\} \\
R_1 &= \max_{x_2\in\mathcal{X}_2} \min\{I(X;Y_1|X_1,x_2), I(X;Y|X_1,x_2)+I(X_1;Y|x_2)\} \\
R_2 &= \max_{x_1\in\mathcal{X}_1} \min\{I(X;Y_2|X_2,x_1), I(X;Y|X_2,x_1)+I(X_2;Y|x_1)\} \\
R_{12} &= \min\{I(X;Y_1|X_1,X_2), I(X;Y_2|X_1,X_2)+I(X_1;Y_2|X_2), I(X;Y|X_1,X_2)+I(X_1;Y|X_2)+I(X_2;Y)\} \\
R_{21} &= \min\{I(X;Y_2|X_1,X_2), I(X;Y_1|X_1,X_2)+I(X_2;Y_1|X_1), I(X;Y|X_1,X_2)+I(X_2;Y|X_1)+I(X_1;Y)\} \\
R_G &= \min\{I(X;Y_1|X_1,X_2), I(X;Y_2|X_1,X_2), I(X,X_1,X_2;Y)\},
\end{aligned}$$

where R_1 is the rate obtained when only relay 1 is active, R_2 is the rate obtained when only relay 2 is active, R_{12} is the rate obtained when relay 1 decodes first and relay 2 decodes second and R_{21} is the rate obtained when this order is reversed. R_G is the rate obtained when both relays form one group[2]. Now, as in the single-relay case, DAF is limited by the worst source-relay link. Therefore, if

$$R^{PTP} > \max_{p(x|x_1,x_2),(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} \big\{I(X;Y_1|x_1,x_2), I(X;Y_2|x_1,x_2)\big\}, \tag{34}$$

where $R^{PTP} = \max_{p(x|x_1,x_2),(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2} I(X;Y|x_1,x_2)$ is the point-to-point rate, then it is better not to use [5, theorem 3.1] at all, but rather set the relays to transmit the symbol pair $(x_1,x_2)\in\mathcal{X}_1\times\mathcal{X}_2$ such that the point-to-point rate is maximized. However, the rate obtained using corollary 2 for the two-relay case is given by

$$\begin{aligned}
R^{TS-EAF} \le \sup_{p(x)p(x_1)p(x_2)} & I(X;Y|X_1,X_2) + q_1(1-q_2)I(X;Y_1|X_1,X_2,Y) \\
& + (1-q_1)q_2 I(X;Y_2|X_1,X_2,Y) + q_1q_2 I(X;Y_1,Y_2|X_1,X_2,Y),
\end{aligned}$$

where q_1 and q_2 are positive and determined according to (28). This expression can, in general be greater than R^{PTP} even when (34) holds, for channels where the relay to destination links are very good. Hence, this explicit achievable expression provides an easy way to improve upon the DAF-based achievable rates when the source-to-relay links are very noisy.

To demonstrate this, consider the channel given in table I over binary RVs X, X_1, X_2, Y, Y_1 and Y_2. The channel distribution was constructed under the independence constraint

$$p(y,y_1,y_2|x,x_1,x_2) = p(y_1|x,x_1,x_2)p(y_2|x,x_1,x_2)p(y|x,x_1,x_2,y_1,y_2),$$

i.e. given the channel inputs, the two relay outputs are independent. This channel is characterized by noisy source-relay links, while the link from relay 1 to the destination has low noise. Therefore, DAF is inferior to the point-to-point transmission but EAF is able to exceed this rate, by giving up a small amount of rate on the direct link (compared to the point-to-point rate) and gaining more rate through the relays. The numerical evaluation of the

[2] In fact, since we take the supremum over all p.d.f.'s $p(x,x_1,x_2)$ we do not need to explicitly include R_1 and R_2 in the maximization, but it is included here to provide a complete presentation.

TABLE I

$p(y, y_1, y_2|x, x_1, x_2)$ FOR THE EAF EXAMPLE.

(x, x_1, x_2)	$p(y, y_1, y_2\|x, x_1, x_2)$							
	000	001	010	011	100	101	110	111
000	8.047314e-2	1.948360e-1	2.041506e-1	4.523933e-2	2.423322e-1	7.057734e-3	1.310053e-1	9.490483e-2
001	8.601616e-1	6.643713e-2	1.662897e-2	1.937227e-2	1.859104e-2	1.741020e-2	8.833169e-4	5.154431e-4
010	3.131504e-1	1.821840e-1	5.618147e-2	1.522841e-1	5.290856e-2	1.555570e-1	3.214581e-2	5.558854e-2
011	5.183921e-3	3.704625e-1	1.641795e-2	2.208356e-1	1.660775e-3	2.355928e-1	9.590170e-4	1.488874e-1
100	8.116746e-3	8.139504e-3	9.387860e-2	1.736515e-2	1.039350e-1	7.308714e-3	7.612555e-1	7.612563e-7
101	4.824126e-2	1.196128e-1	1.705739e-1	7.127199e-2	4.631349e-2	1.955324e-1	1.928693e-1	1.555848e-1
110	9.367321e-2	1.248830e-1	1.873302e-1	6.161358e-2	5.827773e-2	1.906660e-1	1.589616e-1	1.245946e-1
111	9.141272e-7	9.141263e-1	7.618061e-3	3.435473e-2	7.974830e-4	4.117531e-2	9.302643e-4	9.969457e-4

TABLE II

OPTIMAL DISTRIBUTION FOR DAF

(x, x_1, x_2)	$p(x, x_1, x_2)$
000	5.698189907239905e-009
001	5.259061814752764e-017
010	4.301809992760095e-009
011	4.424193267301109e-001
100	6.792096128437060e-009
101	4.740938235494830e-017
110	3.207903771562940e-009
111	5.575806532698892e-001

TABLE III

OPTIMAL DISTRIBUTION FOR EAF

$\Pr(X = 0) = 4.3752093552645e-001$
$\Pr(X_1 = 0) = 1.9388669163312e-001$
$\Pr(X_2 = 0) = 1.000000000000000e-009$

rates for this channel produces[3]

$$
\begin{aligned}
R^{PTP} &= 0.2860323, \\
R^{DAF} &= 0.2408629, \\
R^{TS-EAF} &= 0.2924798,
\end{aligned}
$$

where the optimal distributions that achieve these rates are summarized in tables II and III. The optimal DAF distribution fixes both X_1 and X_2 to $'1'$ and sets the probability of X to be $\Pr(X = 1) = 0.442419$, as expected for the case where the relays limit the achievable rate. For the EAF, the useless relay 2 is fixed to 0, to facilitate transmission with the useful relay 1. In accordance, we obtain time sharing proportions of $q_1 = 0.156947$ and $q_2 \approx 0$ for relay 1 and relay 2 respectively. We note that in this scenario, we actually have that even the single-relay TS-EAF outperforms the two-relay DAF.

[3]The resulting rates were obtained by optimizing for the rates with random initial input distributions. The optimization was repeated 50 times for each rate and the maximum resulting rate was recorded. The m-files used for this evaluation are available at `http://cn.ece.cornell.edu`.

IV. THE GAUSSIAN RELAY CHANNEL

In this section we investigate the application of estimate-and-forward with time-sharing to the Gaussian relay channel. For this channel, the common practice it to use Gaussian codebooks and Gaussian quantization at the relay. The rate in Gaussian scenarios where coded modulation is applied, is usually analyzed by applying DAF at the relay. In this section we show that when considering coded modulation, one should select the relay strategy according to the channel condition: Gaussian selection seems a good choice when the SNR at the relay is low and DAF appears to be superior when the relay enjoys high SNR conditions. However, for intermediate SNR there is much room for optimizing the estimation mapping at the relay.

In the following we first recall the Gaussian relay channel with a Gaussian codebook, and then we consider the Gaussian relay channel under BPSK modulation constraint. Since we focus on the mapping at the relay we consider here the Gaussian relay channel with an orthogonal relay of finite capacity C, also considered in [11]. This scenario is depicted in figure 3.

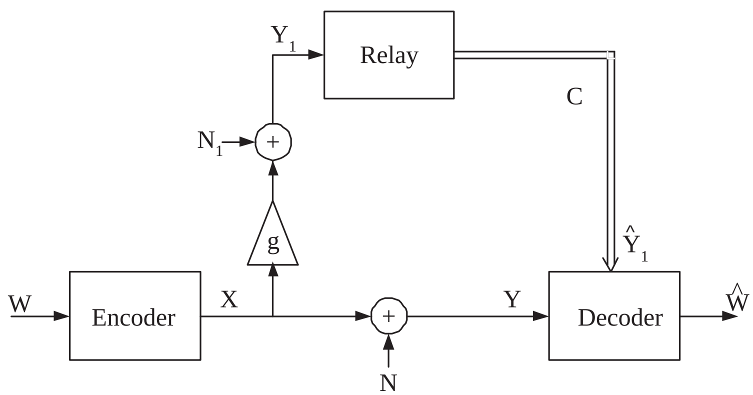

Fig. 3. The Gaussian relay channel with a finite capacity noiseless relay link between the relay and the destination.

Here $Y_1 = g \cdot X + N_1$ is the channel output at the relay, $Y = X + N$ is the channel output at the receiver, which decodes the message based on $(Y^n, \hat{Y}_1^n)$. Let $\mathcal{W} = \left\{1, 2, ..., 2^{nR}\right\}$ denote the source message set, and let the source have an average power constraint P:

$$\frac{1}{n}\sum_{i=1}^{n} x_i(w) \leq P, \qquad \forall w \in \mathcal{W}.$$

The relay signal $\hat{Y}_1^n$ is transmitted to the destination through a finite-capacity noiseless link of capacity C. For this scenario the expressions of [2, theorem 6] specialize to

$$R \leq I(X; Y, \hat{Y}_1) \tag{35a}$$

$$\text{subject to } C \geq I(\hat{Y}_1; Y_1|Y), \tag{35b}$$

with the Markov chain $X, Y - Y_1 - \hat{Y}_1$.

We also consider in this section the DAF method whose information rate is given by (see [2, theorem 1])

$$R_{DAF} = \min\left\{I(X;Y_1), I(X;Y)+C\right\},$$

and the upper bound of [2, theorem 3]:

$$R_{upper} = \min\left\{I(X;Y)+C, I(X;Y,Y_1)\right\}.$$

We note that although these expressions were derived for the finite, discrete alphabets case, following the argument in [8, remark 30], they also hold for the Gaussian case.

A. *The Gaussian Relay Channel with Gaussian Codebooks*

When $X \sim \mathcal{N}(0,P)$, i.i.d., then the channel outputs at the relay and the receiver are jointly Normal RVs:

$$\begin{pmatrix} y \\ y_1 \end{pmatrix} \sim \mathcal{N}\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} P+\sigma^2 & gP \\ gP & g^2P+\sigma_1^2 \end{pmatrix}\right).$$

The compression is achieved by adding to Y_1 a zero mean independent Gaussian RV, N_Q:

$$\hat{Y}_1 = Y_1 + N_Q, \qquad N_Q \sim \mathcal{N}(0,\sigma_Q^2). \tag{36}$$

We refer to the assignment (36) as Gaussian-quantization estimate-and-forward (GQ-EAF). Evaluating the expressions (35a) and (35b) with assignment (36) results in (see also [11]):

$$I(X;Y,\hat{Y}_1) = \frac{1}{2}\log_2\left(1+P+\frac{gP}{1+\sigma_Q^2}\right) \tag{37a}$$

$$I(Y_1;\hat{Y}_1|Y) = \frac{1}{2}\log_2\left(1+\frac{1+P+gP}{\sigma_Q^2(P+1)}\right). \tag{37b}$$

The feasibility condition (35b) yields

$$\sigma_Q^2 \geq \frac{1+P+gP}{(2^{2C}-1)(P+1)},$$

and because maximizing the rate (37a) requires minimizing σ_Q^2, the resulting GQ-EAF rate expression is

$$R \leq \frac{1}{2}\log_2\left(1+P+\frac{gP}{1+\frac{1+P+gP}{(2^{2C}-1)(P+1)}}\right).$$

Now, when using Gaussian quantization at the relay it is obvious that time sharing does not help: we need the minimum σ_Q^2 in order to maximize the rate. This minimum is obtained only when the entire capacity of the relay link is dedicated to the transmission of the (minimally) quantized Y_1. However, when we consider the Gaussian relay channel with coded modulation, the situation is quite different, as we show in the remaining of this section.

B. The Gaussian Relay Channel with Coded Modulation

Consider the Gaussian relay channel where X is an equiprobable BPSK signal of amplitude $\sqrt{P}$:

$$\Pr(X=\sqrt{P})=\Pr(X=-\sqrt{P})=\frac{1}{2}. \tag{38}$$

Under these conditions, the received symbols (Y,Y_1) are no longer jointly Gaussian, but follow a Gaussian-mixture distribution:

$$\begin{aligned} f(y,y_1) &= \Pr(X=\sqrt{P})f(y,y_1|x=\sqrt{P})+\Pr(X=-\sqrt{P})f(y,y_1|x=-\sqrt{P}) \\ &= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2)G_{y_1}(g\sqrt{P},\sigma_1^2)+G_y(-\sqrt{P},\sigma^2)G_{y_1}(-g\sqrt{P},\sigma_1^2)\right), \end{aligned}$$

where

$$G_x(a,b)\triangleq\frac{1}{\sqrt{2\pi b}}e^{-\frac{(x-a)^2}{2b}}. \tag{39}$$

Contrary to the Gaussian codebook case, where it is hard to identify a mapping $p(\hat{y}_1|y_1)$ that will be superior to Gaussian quantization (if indeed such a mapping exists), in this case it is a natural question to compare the Gaussian mapping of (36), which induces a Gaussian-mixture distribution for $\hat{Y}_1$ with other possible mappings. In the case of binary inputs it is natural to consider binary mappings for $\hat{Y}_1$. We can predict that such mappings will do well at high SNR on the source-relay link, when the probability of error for symbol-by-symbol detection at the relay is small, with a much smaller complexity than Gaussian quantization. We start by considering two types of hard-decision (HD) mappings:

1) The first mapping is HD-EAF: The relay first makes a hard decision about every received Y_1 symbol, determining whether it is positive or negative, and then randomly decides if it is going to transmit this decision or transmit an erasure symbol E instead. The probability of transmitting an erasure, $1-P_{\text{no erase}}$, is used to adjust the conference rate such that the feasibility constraint is satisfied. Therefore, the conditional distribution $p(\hat{Y}_1|Y_1)$ is given by:

$$p(\hat{Y}_1|Y_1>0)=\begin{cases} P_{\text{no erase}} & ,1 \\ 1-P_{\text{no erase}} & ,E \end{cases} \tag{40a}$$

$$p(\hat{Y}_1|Y_1\leq 0)=\begin{cases} P_{\text{no erase}} & ,-1 \\ 1-P_{\text{no erase}} & ,E \end{cases}. \tag{40b}$$

This choice is motivated by the time-sharing method considered in section II: after making a hard decision on the received symbol's sign — positive or negative, the relay applies TS to that decision so that the rate required to transmit the resulting random variable is less than C. This facilitates transmission to the destination through the conference link. Since the entropy of the sign decision is 1, then when $C\geq 1$ we can transmit the sign decisions directly without using an erasure. Therefore, we expect that for values of C in the range $C>1$, this mapping will not exceed the rate obtained for $C=1$. The focus is, therefore, on values of C that are less than 1. The expressions for this assignment are given in appendix A-A.

2) The second method is deterministic hard-decision. In this approach, we select a threshold T such that the range of Y_1 is partitioned into three regions: $Y_1 < -T, -T \leq Y_1 \leq T, Y_1 > T$. Then, according to the value of each received Y_1 symbol, the corresponding $\hat{Y}_1$ is deterministically determined:

$$\hat{Y}_1 = \begin{cases} 1, & Y_1 > T \\ E, & -T \leq Y_1 \leq T \\ -1, & Y_1 < -T \end{cases} . \tag{41}$$

The threshold T is selected such that the achievable rate is maximized subject to satisfying the feasibility constraint. We refer to this method as deterministic HD (DHD). Therefore, this is another type of TS in which the erasure probability is determined by the fraction of the time the relay input is between $-T$ to T. This method should be better than HD-EAF at high relay SNR since for HD-EAF, erasure is selected without any regard to the quality of the decision - both good sign decisions and bad sign decisions are erased with the same probability. However in DHD, the erased area is the area where the decisions have low quality in the first place and all high quality decisions are sent. However, at low relay SNR and small capacity for the relay-destination link, HD-EAF may perform better than DHD since the erased area (i.e. the region between $-T$ to $+T$) for the DHD mapping has to be very large to allow 'squeezing' the estimate through the relay link, while HD-EAF may require less compression of the HD output. The expressions for evaluating the rate of the DHD assignment are given in appendix A-B.

We now examine the performance of each technique using numerical evaluation: first, we examine the achievable rates with HD-EAF. The expressions are evaluated for $\sigma_1^2 = \sigma^2 = 1$ and $P = 1$. For every pair of values (g, C) considered, the maximum $P_{\text{no erase}}$ was selected. Figure 4 depicts the achievable rate vs. g for $0.4 \leq C \leq 2$, together with the upper bound and the decode-and-forward rate. As can be observed from figure 4, the information rate of HD-EAF increases with C until $C = 1$ and then remains constant. It is also seen that for small values of g, HD-EAF is better than DAF. This region of g increases with C, and for $C \geq 1$ the crossover value of g is approximately 1.71. However, even for $g = 2$, DAF is only 2.5% better than HD-EAF.

Next, examine DHD: as can be seen from figure 5, for small values of C, DAF exceeds the information rate of DHD for values of g greater than 1, but for $C \geq 0.8$, DHD is superior to DAF, and in fact DAF approaches DHD from below. Another phenomena obvious from the figure (esp. for $C = 0.8$), is the existence of a threshold: for low values of C there is some g at which the DHD rate exhibits a jump. This can be explained by looking at figure 6, which depicts the values of $I(X; \hat{Y}_1, Y)$ and $I(\hat{Y}_1; Y_1|Y)$ vs. the threshold T: the bold-solid graph of $I(\hat{Y}_1; Y_1|Y)$ can intersect the bold-dashed horizontal line representing C at two values of T. We also note that for small T the value of $I(X; \hat{Y}_1, Y)$ is generally greater than for large T. Now, the jump can be explained as follows: as shown in appendix A-B.1, for small T and g, $I(\hat{Y}_1; Y_1|Y)$ is bounded from below. Now, if this bound value is greater than C then the intersection will occur only at a large value of T, hence the small rate. When g increases, the value of $I(\hat{Y}_1; Y_1|Y)$ for small T decreases accordingly, until at some g it intersects C for a small T as well as for a large T, as indicated by the arrow in the right-hand part of figure 6. This allows us to obtain the rates in

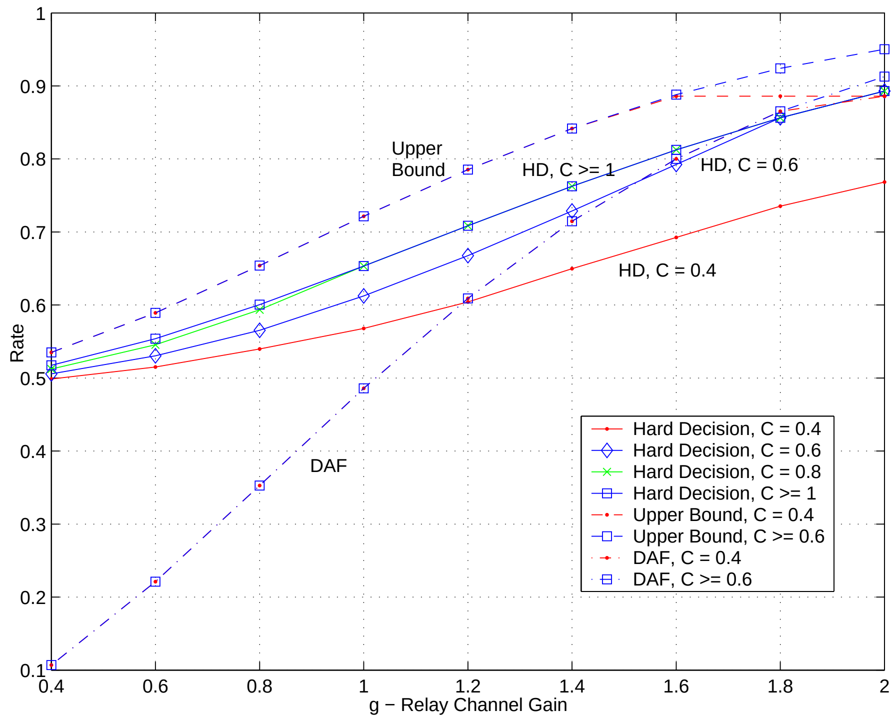

Fig. 4. Information rate with BPSK and hard decision EAF mapping at the relay vs. relay channel gain g, for different values of C.

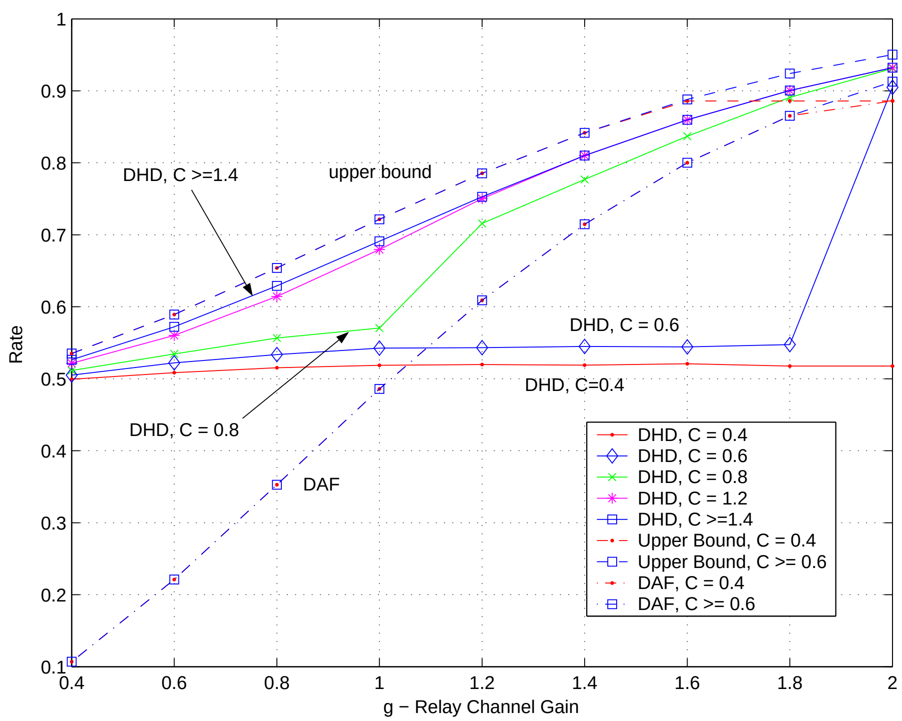

Fig. 5. Information rate with BPSK, for deterministic hard decision at the relay vs. relay channel gain g, for different values of C.

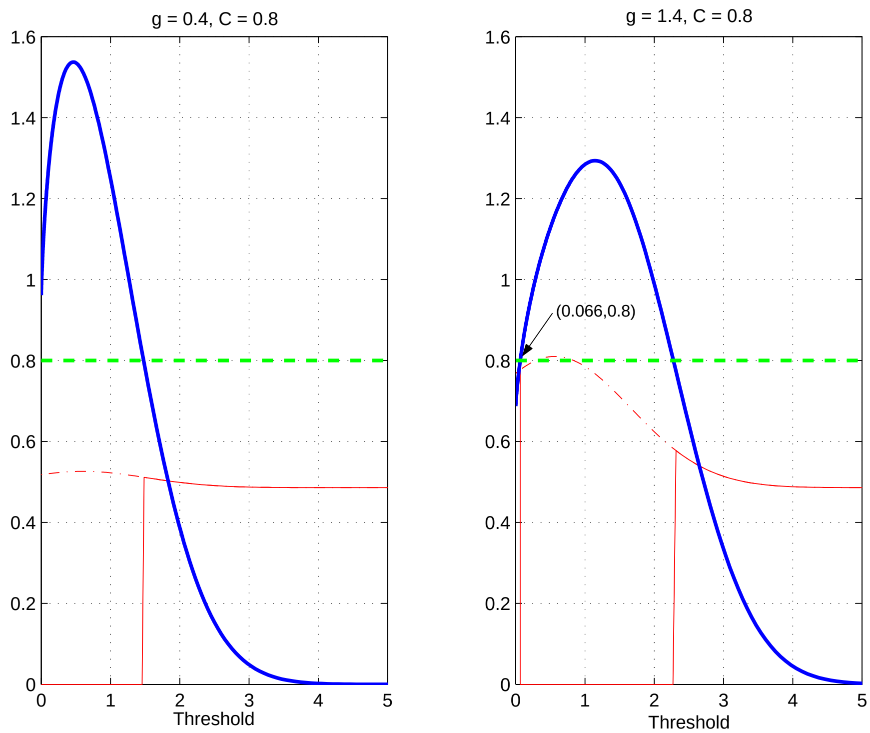

Fig. 6. $I(\hat{Y}_1; Y_1|Y)$ and $I(X; \hat{Y}_1, Y)$ vs. Threshold T for $(g, C) = (0.4, 0.8)$ (left) and $(g, C) = (1.4, 0.8)$ (right). The bold solid line represents $I(\hat{Y}_1, Y_1|Y)$, the bold dashed line represents $C = 0.8$, $I(X; Y, \hat{Y}_1)$ is represented by the dash-dot line and the resulting information rate is depicted with the solid line.

the region of small T which are in general higher than the rates for large T and this is the source of the jump in the achievable rate.

C. Time-Sharing Deterministic Hard-Decision (TS-DHD)

It is clearly evident from the above numerical evaluation that none of the two mappings, HD-EAF and DHD, is universally better than the other: when g is small and C is less than 1, then HD-EAF performs better than DHD, since the erased region is too large, and when g increases, DHD performs better than HD-EAF since it erases only the low quality information. It is therefore natural to consider a third mapping which combines both aspects of binary mapping at the relay, namely deterministically erasing low quality information and then randomly gating the resulting discrete variable in order to allow its transmission over the conference link. This hybrid mapping is

given in the following equation:

$$p(\hat{Y}_1|Y_1 > T) = \begin{cases} P_{\text{no erase}} & ,1 \\ 1 - P_{\text{no erase}} & ,E \end{cases} \tag{42a}$$

$$p(\hat{Y}_1 = E \mid |Y_1| \leq T) = 1 \tag{42b}$$

$$p(\hat{Y}_1|Y_1 < -T) = \begin{cases} P_{\text{no erase}} & ,-1 \\ 1 - P_{\text{no erase}} & ,E \end{cases} . \tag{42c}$$

In this mapping, the region $|Y_1| \leq T$ is always erased, and the complement region is erased with probability $P_{\text{erase}} = 1 - P_{\text{no erase}}$. Of course, now both T and P_{erase} have to be optimized. The expressions for TS-DHD can be found in appendix A-C. Figure 7 compares the performance of DHD, HD-EAF and TS-DHD. As can be seen, the hybrid method enjoys the benefits of both types of mappings and is the superior method.

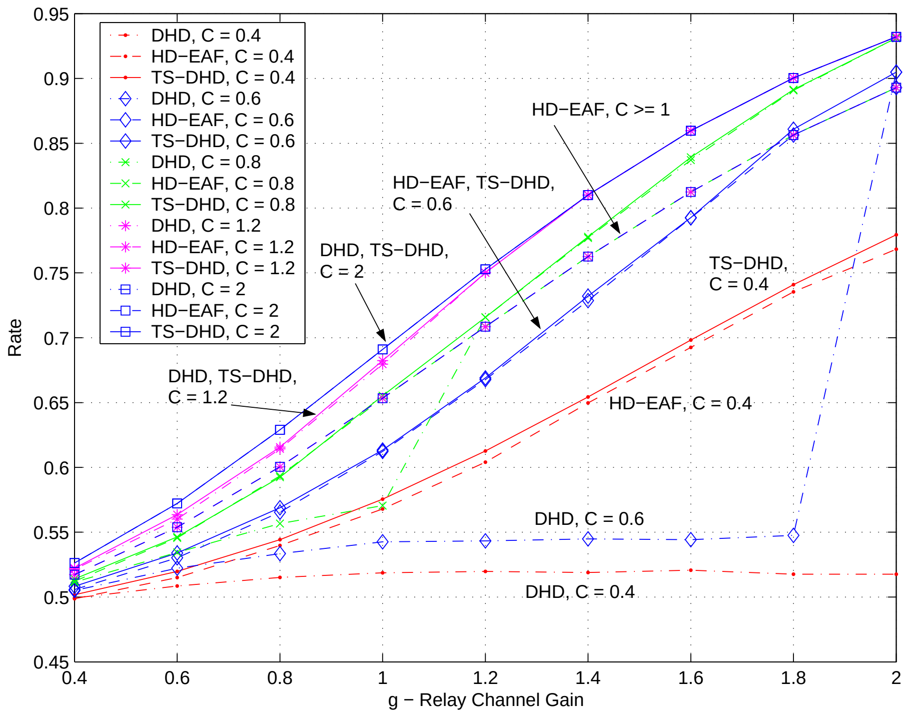

Fig. 7. Information rate with BPSK, for HD-EAF, DHD and TS-DHD at the relay vs. relay channel gain g, for different values of C.

Next, figure 8 compares the performance of TS-DHD, GQ-EAF, and DAF. As can be seen from the figure, Gaussian quantization is not always the optimal choice: for $C = 0.6$ (the lines with diamond-shaped markers) we have that GQ-EAF is the best method for $g < 1.05$, for $1.05 < g < 1.55$ TS-DHD is the best method and for $g > 1.55$ DAF achieves the highest rate. For $C = 1$ (x-shaped markers) TS-DHD is superior to both GQ-EAF and DAF for $g > 0.9$ and for $C = 2$, GQ-EAF is the superior method for all $g \leq 2$. This suggests that for the practical Gaussian relay scenario, where the modulation constraint is taken into account, there is room to optimize the mapping at the relay since the choice of Gaussian quantization is not always optimal.

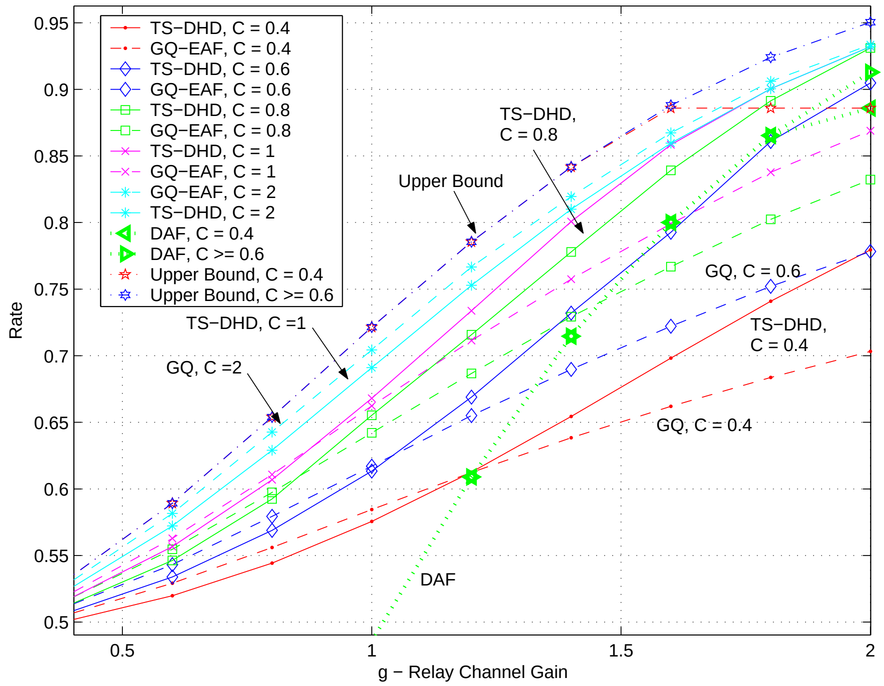

Fig. 8. Information rate with BPSK, for DAF, TS-DHD and GQ-EAF at the relay vs. relay channel gain g, for different values of C.

Lastly, figure 9 depicts the regions in the g-C plane in which each of the methods considered here is superior, in a similar manner to [11, figure 2][4]. As can be observed from the figure, in the noisy region of small g and also in the region of very large C, GQ-EAF is superior, and in the strong relay region of medium-to-high g and medium-to-high C, TS-DHD is the superior method. DAF is superior small C and high g. In a sense, the TS-DHD method is a hybrid method between the DAF which makes a hard-decision on the entire block and GQ-EAF which makes a soft decision every symbol, therefore it is superior in the transition region between the region where DAF is distinctly better, and the region where GQ-EAF is distinctly superior.

[4]The block shapes are due to the step-size of 0.2 in the values of g and C used for evaluating the rates. In the final version we will present an evaluation over a finer grid (such an evaluation requires several weeks to complete).

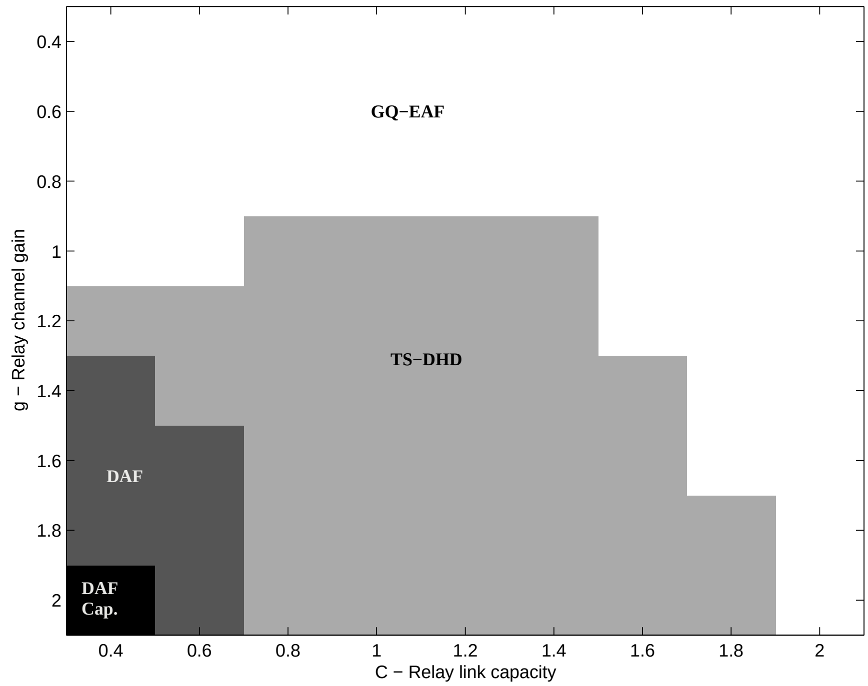

Fig. 9. The best cooperation strategy (out of DAF, TS-DHD and GQ-EAF) for the Gaussian relay channel with BPSK transmission.

D. When the SNR on the Direct Link Approaches 0 *($\sigma^2 \to \infty$)*

In this subsection we analyze the relaying strategies discussed in this section as the SNR on the direct link $X - Y$ approaches zero. Because TS-DHD is a hybrid method combining both DHD and HD-EAF, we analyze the behavior of the components rather than the hybrid, to gain more insight. This analysis is particularly useful when trying to numerically evaluate the rates, since as the direct-link SNR goes to zero, the computer's numerical accuracy does not allow to numerically obtain the rates using the general expressions.

First we note that when the SNR of the direct link $X - Y$ approaches 0 we have that $I(X;Y) \to 0$ as well. To see this we write

$$\begin{aligned} I(X;Y) &= h(Y) - h(Y|X) \\ &= h(Y) - h(X+N|X) \\ &= h(Y) - h(N), \end{aligned}$$

with $h(Y) = -\int_{-\infty}^{\infty} f(y)\log_2(f(y))dy$, and from (A.3)

$$\begin{aligned}
f(Y) &= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2) + G_y(-\sqrt{P},\sigma^2)\right)\\
&= \frac{1}{2}\left(\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y-\sqrt{P})^2}{2\sigma^2}} + \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(y+\sqrt{P})^2}{2\sigma^2}}\right)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}e^{\frac{y\sqrt{P}}{\sigma^2}} + \frac{1}{2}e^{-\frac{y\sqrt{P}}{\sigma^2}}\right)e^{-\frac{P}{2\sigma^2}}\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right)e^{-\frac{P}{2\sigma^2}}\\
&\overset{\sigma^2\to\infty}{\approx} \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\\
&\triangleq G_y(0,\sigma^2),
\end{aligned}$$

where the approximation is in the sense that for small $|y|$ we have $\cosh(|y|) \approx 1$ and for large $|y|$, $e^{-\frac{y^2}{2\sigma^2}}$ drives the entire expression to zero as $e^{-\frac{y^2}{2\sigma^2}}$, for $\sigma^2 \to \infty$. This approximation reflects the intuitive notion that as the variance increases to infinity, the two-component, symmetric Gaussian mixture resembles more and more a zero-mean Gaussian RV with the same variance. Therefore, for low SNR, the output is very close to a zero-mean Normal RV with variance σ^2, and $h(Y) \approx h(N)$,[5] hence

$$I(X;Y) \overset{\sigma^2\to\infty}{\longrightarrow} 0.$$

Note that the upper bound and the decode-and-forward rate in this case are both equal to

$$R_{DAF} = R_{upper} = \min\left\{C, I(X;Y_1)\right\}.$$

Now, let us evaluate the rate for HD-EAF as the SNR goes to zero. From (35a):

$$R \leq I(X;Y,\hat{Y}_1) = I(X;\hat{Y}_1) + I(X;Y|\hat{Y}_1),$$

and

$$\begin{aligned}
I(X;Y|\hat{Y}_1) &= h(Y|\hat{Y}_1) - h(Y|X,\hat{Y}_1)\\
&= \Pr(\hat{Y}_1 = 1)h(Y|\hat{Y}_1 = 1) + \Pr(\hat{Y}_1 = E)h(Y|\hat{Y}_1 = E) + \Pr(\hat{Y}_1 = -1)h(Y|\hat{Y}_1 = -1) - h(N).
\end{aligned}$$

[5]For $\sigma = 20$ we have that $\int_{-\infty}^{\infty} |f_Y(y) - G_y(0,\sigma^2)|dy < 0.001$, for $\sigma = 55$, $h(Y) - h(N) \approx 0.001$ and for $\sigma = 200$, $h(Y) - h(N) < 0.0001$.

Using appendix A, equations (A.5) – (A.7), we have

$$\begin{aligned}
h(Y|\hat{Y}_1=1) &= -\int_{y=-\infty}^{\infty} f_{Y|\hat{Y}_1}(y|\hat{y}_1=1)\log_2\left(f_{Y|\hat{Y}_1}(y|\hat{y}_1=1)\right)dy,\\
f_{Y|\hat{Y}_1}(y|\hat{y}_1=1) &= \frac{f_{Y,Y_1}(y,y_1>0)P_{\text{no erase}}}{\Pr(Y_1>0)P_{\text{no erase}}} = \frac{f_{Y,Y_1}(y,y_1>0)}{\Pr(Y_1>0)},\\
f_{Y,Y_1}(y,y_1>0) &= \frac{1}{2}\left(f_{Y,Y_1|X}(y,y_1>0|x=\sqrt{P}) + f_{Y,Y_1|X}(y,y_1>0|x=-\sqrt{P})\right)\\
&= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2)\Pr(Y_1>0|X=\sqrt{P}) + G_y(-\sqrt{P},\sigma^2)\big(1-\Pr(Y_1>0|X=\sqrt{P})\big)\right)\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}e^{\frac{y\sqrt{P}}{\sigma^2}}\Pr(Y_1>0|X=\sqrt{P}) + \frac{1}{2}e^{-\frac{y\sqrt{P}}{\sigma^2}}\big(1-\Pr(Y_1>0|X=\sqrt{P})\big)\right)e^{-\frac{P}{2\sigma^2}}\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{\left(\frac{1}{2}-\delta\right)e^{\frac{y\sqrt{P}}{\sigma^2}} + \left(\frac{1}{2}+\delta\right)e^{-\frac{y\sqrt{P}}{\sigma^2}}}{2}\right)e^{-\frac{P}{2\sigma^2}}\\
&= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right) - \delta\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right)\right)e^{-\frac{P}{2\sigma^2}}\\
&\overset{(a)}{\approx} \frac{1}{2}G_y(0,\sigma^2),
\end{aligned}$$

when $\sigma^2 \to \infty$ and $\delta \in \left[-\frac{1}{2},\frac{1}{2}\right]$ is selected such that $\Pr(Y_1>0|X=\sqrt{P}) = \frac{1}{2}-\delta$. The approximation in (a) is because for small $|y|$, $\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \approx 0$ and $\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \approx 1$, and for large $|y|$, both $e^{-\frac{y^2}{2\sigma^2}}\sinh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \to 0$ and $e^{-\frac{y^2}{2\sigma^2}}\cosh\left(\frac{y\sqrt{P}}{\sigma^2}\right) \to 0$. Hence

$$\begin{aligned}
h(Y|\hat{Y}_1=1) &\approx -\int_{y=-\infty}^{\infty}\frac{G_y(0,\sigma^2)}{2\Pr(Y_1>0)}\log_2\left(\frac{G_y(0,\sigma^2)}{2\Pr(Y_1>0)}\right)dy\\
&= -\frac{1}{2\Pr(Y_1>0)}\int_{y=-\infty}^{\infty}G_y(0,\sigma^2)\left[\log_2\left(G_y(0,\sigma^2)\right) - \log_2\left(2\Pr(Y_1>0)\right)\right]dy\\
&= \frac{1}{2\Pr(Y_1>0)}\left[h(N) + \log_2\left(2\Pr(Y_1>0)\right)\right],
\end{aligned}$$

and using $\Pr(Y_1>0) = \Pr(Y_1\le 0) = \frac{1}{2}$ and $h(Y|\hat{Y}_1=1) = h(Y|\hat{Y}_1=-1)$, we obtain

$$\begin{aligned}
h(Y|\hat{Y}_1) &\approx \frac{1}{2}P_{\text{no erase}}h(N) + (1-P_{\text{no erase}})h(N) + \frac{1}{2}P_{\text{no erase}}h(N)\\
&= h(N).
\end{aligned}$$

Therefore, at low SNR, Y and $\hat{Y}_1$ become independent. Then, $I(X;Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(N) \approx 0$ and the information rate becomes (see appendix A-E)

$$\begin{aligned}
R \le I(X;\hat{Y}_1) &= H(\hat{Y}_1) - H(\hat{Y}_1|X)\\
&= P_{\text{no erase}}(1 - H(P_1, 1-P_1)),
\end{aligned}$$

where $H(\cdot)$ is the discrete entropy for the specified discrete distribution and $P_1 = \Pr(Y_1>0|X=\sqrt{P})$. Now,

consider the feasibility condition $C \geq I(Y_1; \hat{Y}_1|Y)$:

$$\begin{aligned} I(Y_1; \hat{Y}_1|Y) &= H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y_1, Y) \\ &\overset{(a)}{\approx} H(\hat{Y}_1) - H(\hat{Y}_1|Y_1) \\ &= P_{\text{no erase}}, \end{aligned}$$

where (a) follows from the independence of Y and $\hat{Y}_1$ at low SNR, see appendix A-E. Therefore, for low SNR, we set $P_{\text{no erase}} = \min\{C, 1\}$ and the rate becomes

$$R \leq \min\{C, 1\}\,(1 - H(P_1, 1 - P_1)).$$

For the GQ-EAF we first approximate $f(Y, \hat{Y}_1)$ at low SNR starting with (A.8):

$$\begin{aligned} f_{Y,\hat{Y}_1}(y, \hat{y}_1) &= \frac{1}{2}\left(G_y(\sqrt{P}, \sigma^2)G_{\hat{y}_1}(g\sqrt{P}, \sigma_1^2 + \sigma_Q^2) + G_y(-\sqrt{P}, \sigma^2)G_{\hat{y}_1}(-g\sqrt{P}, \sigma_1^2 + \sigma_Q^2)\right) \\ &= \frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{y^2}{2\sigma^2}}\left(\frac{1}{2}G_{\hat{y}_1}(g\sqrt{P}, \sigma_1^2 + \sigma_Q^2)e^{\frac{y\sqrt{P}}{\sigma^2}} + \frac{1}{2}G_{\hat{y}_1}(-g\sqrt{P}, \sigma_1^2 + \sigma_Q^2)e^{\frac{-y\sqrt{P}}{\sigma^2}}\right)e^{-\frac{P}{2\sigma^2}} \\ &\approx G_y(0, \sigma^2)f_{\hat{Y}_1}(\hat{y}_1), \end{aligned}$$

as $e^{\pm\frac{y\sqrt{P}}{\sigma^2}} \approx 1$ in the region when $G_{\hat{y}_1}$ is significant, for both $X = \sqrt{P}$ or $X = -\sqrt{P}$. We conclude that as the direct SNR approaches 0, Y and $\hat{Y}_1$ become independent. Now, the rate is given by:

$$\begin{aligned} R &\leq I(X; Y, \hat{Y}_1) \\ &= h(Y, \hat{Y}_1) - h(Y, \hat{Y}_1|X) \\ &= h(Y) + h(\hat{Y}_1) - h(X + N, gX + N_1 + N_Q|X) \\ &= h(Y) + h(\hat{Y}_1) - h(N, N_1 + N_Q|X) \\ &= h(Y) - h(N|X) + h(\hat{Y}_1) - h(N_1 + N_Q|X) \\ &= I(X; Y) + I(X; \hat{Y}_1) \\ &\approx I(X; \hat{Y}_1) \\ &= h(\hat{Y}_1) - h(N_1 + N_Q). \end{aligned} \tag{43}$$

The feasibility condition becomes:

$$\begin{aligned} C &\geq I(\hat{Y}_1; Y_1|Y) \\ &= h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y, Y_1) \\ &\approx h(\hat{Y}_1) - h(N_Q), \end{aligned} \tag{44}$$

with

$$f_{\hat{Y}_1}(\hat{y}_1) = \frac{1}{2}\left[G_{\hat{y}_1}(g\sqrt{P}, \sigma_1^2 + \sigma_Q^2) + G_{\hat{y}_1}(-g\sqrt{P}, \sigma_1^2 + \sigma_Q^2)\right].$$

For DHD, as $\sigma^2 \to \infty$ we have

$$\begin{aligned} I(X;\hat{Y}_1;Y) &= I(X;Y) + I(X;\hat{Y}_1|Y) \\ &\approx I(X;\hat{Y}_1|Y) \\ &= H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y,X) \\ &\overset{(a)}{\approx} H(\hat{Y}_1) - H(\hat{Y}_1|X) \\ &= I(X;\hat{Y}_1) \end{aligned}$$

where (a) follows from the independence of Y and Y_1 as $\sigma^2 \to \infty$ and the fact that $\hat{Y}_1$ is a deterministic function of Y_1, combined with the fact that given X, Y_1 and Y are independent. The feasibility condition becomes

$$C \geq H(\hat{Y}_1|Y) \approx H(\hat{Y}_1).$$

Because $I(X;\hat{Y}_1)$ is not a monotone function of T we have to optimize over T to find the actual rate.

As can be seen from the expression for HD-EAF, when the SNR on the direct link decreases, the capacity of the conference link acts as a scaling factor on the rate of the binary channel from the source to the relay. In figure 10

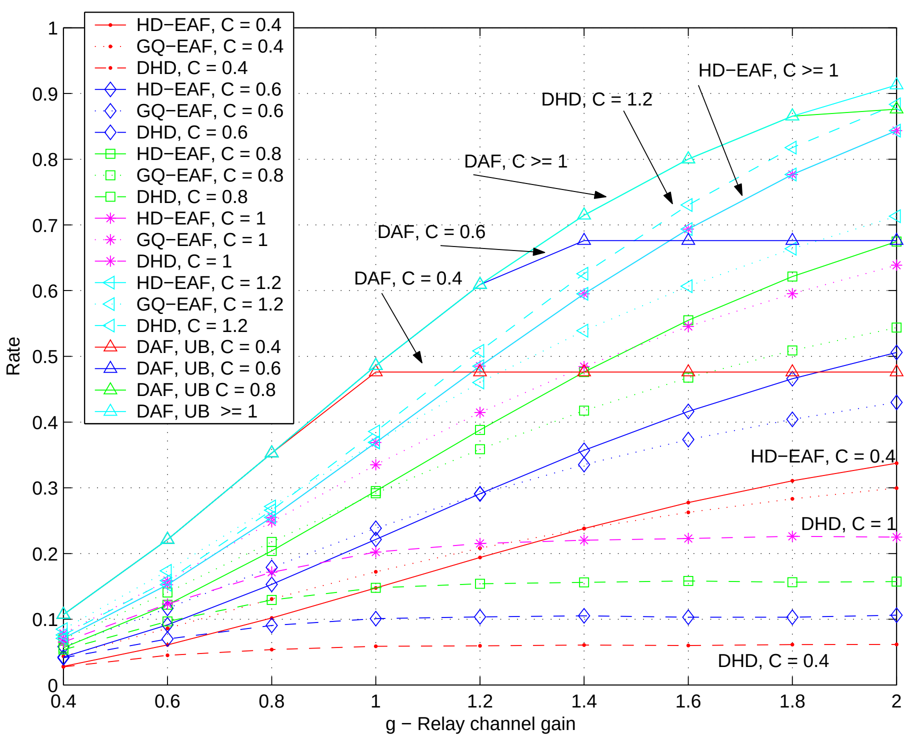

Fig. 10. Information rate with DAF, DHD, HD-EAF and GQ-EAF vs. relay channel gain g, for different values of C, at low SNR on the source-relay link.

we plotted the information rate for DHD, HD-EAF, GQ-EAF and DAF (which coincides with the upper bound). Comparing the three EAF strategies we note that DHD, which at intermediate SNR on the source-relay channel

performs well for $C \geq 0.8$, has the worst performance at low SNR up to $C = 1.2$. At $C = 1.2$, DHD becomes the best technique out of the three. For $C < 1.2$ and high SNR on the source-relay channel, HD-EAF outperforms both DHD and GQ-EAF. For low SNR on the source-relay channel, GQ-EAF is again superior.

E. Discussion

We make the following observations:

- As noted at the beginning of this section, for low SNR on the source-relay link, GQ-EAF outperforms TS-DHD. To see why, consider the distribution of Y_1:

$$
\begin{aligned}
f_{Y_1}(y_1) &= G_{y_1}(0, \sigma_1^2) \cosh\left(\frac{g\sqrt{P}y_1}{\sigma_1^2}\right) e^{-\frac{g^2 P}{2\sigma_1^2}} \\
&\overset{g \to 0}{\approx} G_{y_1}(0, \sigma_1^2) \left(1 - \frac{g^2 P}{2\sigma_1^2}\right),
\end{aligned}
$$

  where the approximation is obtained using the first order Taylor expansion, and the fact that for large values of Y_1, $G_{y_1}(0, \sigma_1^2)$ dominates the expression. Therefore, as $g \to 0$, Y_1 approaches a zero-mean Gaussian RV: $Y_1 \xrightarrow{\mathcal{D}} \mathcal{N}(0, \sigma_1^2)$. As discussed in [24, ch. 13.1], the closer the reconstruction variable is to the original variable, the better the quantization performance are expected to be. Therefore it should be natural to guess that GQ will perform better at low relay link SNR.
- At the other extreme, as $g \to \infty$, consider the DAF strategy: as $g \to \infty$, have that

$$
\begin{aligned}
h(Y_1) &= -\int_{y_1=-\infty}^{\infty} \frac{1}{2}\left[G_{y_1}(g\sqrt{P}, \sigma_1^2) + G_{y_1}(-g\sqrt{P}, \sigma_1^2)\right] \times \\
&\qquad \log_2\left(\frac{1}{2}\left[G_{y_1}(g\sqrt{P}, \sigma_1^2) + G_{y_1}(-g\sqrt{P}, \sigma_1^2)\right]\right) dy_1 \\
&\overset{g \to \infty}{\approx} 1 - \int_{y_1=-\infty}^{\infty} \frac{1}{2} G_{y_1}(g\sqrt{P}, \sigma_1^2) \log_2 G_{y_1}(g\sqrt{P}, \sigma_1^2) dy_1 \\
&\qquad - \int_{y_1=-\infty}^{\infty} \frac{1}{2} G_{y_1}(-g\sqrt{P}, \sigma_1^2) \log_2 G_{y_1}(-g\sqrt{P}, \sigma_1^2) dy_1 \\
&= 1 + h(N_1),
\end{aligned}
$$

  and therefore,

$$
I(X; Y_1) = h(Y_1) - h(Y_1|X) \approx 1 + h(N_1) - h(N_1) = 1 = H(X).
$$

  Hence,

$$
R_{DAF} = \min\left\{I(X; Y_1), I(X; Y) + C\right\} = \min\left\{1, I(X; Y) + C\right\},
$$

  which is the maximal rate. Therefore, as $g \to \infty$ DAF provides the optimal rate.
- We can expect that at intermediate SNR, methods that balance between the soft-decision per symbol of GQ-EAF and the hard-decision on the entire codeword of DAF, will be superior to both. Furthermore, we believe that as the SNR decreases, increasing the cardinality of $\hat{Y}_1$ accordingly will improve the performance.

V. Multi-Step Cooperative Broadcast Application

In this section we consider the cooperative broadcast (BC) scenario. In this scenario, one transmitter communicates with two receivers. In its most general form, the transmitter sends three independent messages: a common message intended for both receivers and two private messages, one for each receiver, where all three messages are encoded into a single channel codeword X^n. Each receiver gets a noisy version of the codeword, Y_1^n at R_{x1} and Y_2^n at R_{x2}. After reception, the receivers exchange messages in a K-cycle conference over noiseless conference links of finite capacities C_{12} and C_{21}. Each conference message is based on the channel output at each receiver and the conference messages previously received from the other receiver, in a similar manner to the conference defined by Willems in [26] for the cooperative MAC. After conferencing, each receiver decodes its message. This scenario is depicted in figure 11. This setup was studied in [12] for the single common message case over the independent BC (i.e. $p(\mathbf{y}_1,\mathbf{y}_2|\mathbf{x}) = \prod_{i=1}^n p(y_{1,i}|x_i)p(y_{2,i}|x_i)$), and in [13] for the general setup with a single cycle of conferencing.

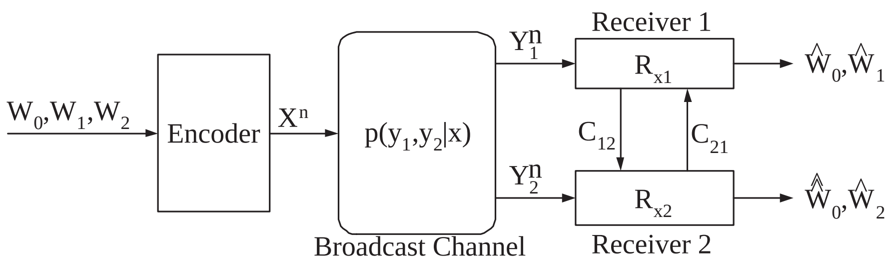

Fig. 11. The broadcast channel with cooperating receivers. The encoder sends three messages, a common message W_0, a private message to R_{x1}, W_1, and a private message to R_{x2}, W_2. $\hat{W}_0$ and $\hat{\hat{W}}_0$ are the estimates of W_0 at R_{x1} and R_{x2} respectively.

A. Definitions

We use the standard definition for the discrete memoryless general broadcast channel given in [28]. We define a cooperative coding scheme as follows:

Definition 5: A (C_{12}, C_{21})-admissible K-cycle conference consists of the following elements:

1) K message sets from R_{x1} to R_{x2}, denoted by $\mathcal{W}_{12}^{(1)}$, $\mathcal{W}_{12}^{(2)}$,...,$\mathcal{W}_{12}^{(K)}$, and K message sets from R_{x2} to R_{x1}, denoted by $\mathcal{W}_{21}^{(1)}$, $\mathcal{W}_{21}^{(2)}$,...,$\mathcal{W}_{21}^{(K)}$. Message set $\mathcal{W}_{12}^{(k)}$ consists of $2^{nR_{12}^{(k)}}$ messages and message set $\mathcal{W}_{21}^{(k)}$ consists of $2^{nR_{21}^{(k)}}$ messages.
2) K mapping functions, one for each conference step from R_{x1} to R_{x2}:
$$h_{12}^{(k)} : \mathcal{Y}_1^n \times \mathcal{W}_{21}^{(1)} \times \mathcal{W}_{21}^{(2)} \times ... \times \mathcal{W}_{21}^{(k-1)} \mapsto \mathcal{W}_{12}^{(k)},$$
and K mapping functions, one for each conference step from R_{x2} to R_{x1}:
$$h_{21}^{(k)} : \mathcal{Y}_2^n \times \mathcal{W}_{12}^{(1)} \times \mathcal{W}_{12}^{(2)} \times ... \times \mathcal{W}_{12}^{(k)} \mapsto \mathcal{W}_{21}^{(k)},$$

where $k = 1, 2, ..., K$.

The conference rates satisfy:

$$C_{12} = \sum_{k=1}^{K} R_{12}^{(k)}, \qquad C_{21} = \sum_{k=1}^{K} R_{21}^{(k)}.$$

Definition 6: A $(2^{nR_0}, 2^{nR_1}, 2^{nR_2}, n, C_{12}, C_{21}, K)$ *code* for the general broadcast channel with a common message and two independent private messages, consists of three sets of source messages, $\mathcal{M}_0 = \left\{1, 2, ..., 2^{nR_0}\right\}$, $\mathcal{M}_1 = \left\{1, 2, ..., 2^{nR_1}\right\}$ and $\mathcal{M}_2 = \left\{1, 2, ..., 2^{nR_2}\right\}$, a mapping function at the transmitter,

$$f : \mathcal{M}_0 \times \mathcal{M}_1 \times \mathcal{M}_2 \mapsto \mathcal{X}^n,$$

A (C_{12}, C_{21})-admissible K-cycle conference, and two decoders,

$$g_1 : \mathcal{W}_{21}^{(1)} \times \mathcal{W}_{21}^{(2)} \times ... \times \mathcal{W}_{21}^{(K)} \times \mathcal{Y}_1^n \mapsto \mathcal{M}_0 \times \mathcal{M}_1,$$

$$g_2 : \mathcal{W}_{12}^{(1)} \times \mathcal{W}_{12}^{(2)} \times ... \times \mathcal{W}_{12}^{(K)} \times \mathcal{Y}_2^n \mapsto \mathcal{M}_0 \times \mathcal{M}_2.$$

Definition 7: The *average probability of error* is defined as the average probability that at least one of the receivers does not decode its message pair correctly:

$$P_e^{(n)} = \Pr\left(g_1\left(W_{21}^{(1)}, W_{21}^{(2)}, ..., W_{21}^{(K)}, Y_1^n\right) \neq (M_0, M_1) \text{ or } g_2\left(W_{12}^{(1)}, W_{12}^{(2)}, ..., W_{12}^{(K)}, Y_2^n\right) \neq (M_0, M_2)\right),$$

where we assume that each message is selected uniformly and independently over its respective message set.

B. The Cooperative Broadcast Channel with Two Independent and One Common Message

We first present the general result for the cooperative broadcast scenario with a K-cycle conference. Denote with $\hat{\mathbf{Y}}_1 = \left(\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}\right)$ and $\hat{\mathbf{Y}}_2 = \left(\hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K)}\right)$. Let R_1 and R_2 be the private rates to R_{x1} and R_{x2} respectively, and let R_0 denote the rate of the common information. Then, the following rate triplets are achievable:

Theorem 4: Consider the general broadcast channel $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ *with cooperating receivers, having noiseless conference links of finite capacities* C_{12} *and* C_{21} *between them. Let the receivers hold a conference that consists of* K *cycles. Then, any rate triplet* (R_0, R_1, R_2) *satisfying*

$$R_0 \leq \min\left\{I\left(W; Y_1, \hat{\mathbf{Y}}_2\right), I\left(W; \hat{\mathbf{Y}}_1, Y_2\right)\right\} \tag{45a}$$

$$R_1 \leq I(U; Y_1, \hat{\mathbf{Y}}_2|W) \tag{45b}$$

$$R_2 \leq I(V; \hat{\mathbf{Y}}_1, Y_2|W) \tag{45c}$$

$$R_1 + R_2 \leq I(U; Y_1, \hat{\mathbf{Y}}_2|W) + I(V; \hat{\mathbf{Y}}_1, Y_2|W) - I(U; V|W), \tag{45d}$$

subject to,

$$C_{12} \geq I(Y_1; \hat{\mathbf{Y}}_1, \hat{\mathbf{Y}}_2|Y_2) \tag{46a}$$

$$C_{21} \geq I(Y_2; \hat{\mathbf{Y}}_2, \hat{\mathbf{Y}}_1|Y_1), \tag{46b}$$

for some joint distribution

$$p\left(w,u,v,x,y_1,y_2,\hat{y}_1^{(1)},\hat{y}_1^{(2)},...,\hat{y}_1^{(K)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},...,\hat{y}_2^{(K)}\right)=$$
$$p(w,u,v,x)p(y_1,y_2|x)p\left(\hat{y}_1^{(1)}|y_1\right)p\left(\hat{y}_2^{(1)}|y_2,\hat{y}_1^{(1)}\right)\cdots p\left(\hat{y}_1^{(k)}|y_1,\hat{y}_1^{(1)},\hat{y}_1^{(2)},...,\hat{y}_1^{(k-1)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},...,\hat{y}_2^{(k-1)}\right)\times$$
$$p\left(\hat{y}_2^{(k)}|y_2,\hat{y}_1^{(1)},\hat{y}_1^{(2)},...,\hat{y}_1^{(k)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},...,\hat{y}_2^{(k-1)}\right)\cdots p\left(\hat{y}_1^{(K)}|y_1,\hat{y}_1^{(1)},\hat{y}_1^{(2)},...,\hat{y}_1^{(K-1)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},...,\hat{y}_2^{(K-1)}\right)$$
$$\times p\left(\hat{y}_2^{(K)}|y_2,\hat{y}_1^{(1)},\hat{y}_1^{(2)},...,\hat{y}_1^{(K)},\hat{y}_2^{(1)},\hat{y}_2^{(2)},...,\hat{y}_2^{(K-1)}\right), \tag{47}$$

is achievable. The cardinality of the k'th auxiliary random variables are bounded by:

$$||\hat{\mathcal{Y}}_1^{(k)}|| \le ||\mathcal{Y}_1||\times\prod_{l=1}^{k-1}||\hat{\mathcal{Y}}_1^{(l)}||\times\prod_{l=1}^{k-1}||\hat{\mathcal{Y}}_2^{(l)}||+1, \qquad k=1,2,...,K$$
$$||\hat{\mathcal{Y}}_2^{(k)}|| \le ||\mathcal{Y}_2||\times\prod_{l=1}^{k}||\hat{\mathcal{Y}}_1^{(l)}||\times\prod_{l=1}^{k-1}||\hat{\mathcal{Y}}_2^{(l)}||+1, \qquad k=1,2,...,K.$$

Proof:

1) Overview of Strategy: The coding strategy is based on combining the BC code construction of [29], after incorporating the common message into the construction, with the K-cycle conference of [30]. The transmitter constructs a broadcast code to split the rate between the three message sets. This is done independently of the relaying scheme. Each receiver generates its conference messages according to the construction of [30]. After K cycles of conferencing each receiver decodes its information based on its channel output and the conference messages received from the other receiver.

2) Code Construction at The Transmitter:

- Fix all the distributions in (47). Fix $\epsilon>0$ and let $n>1$. Let $\delta>0$ be a positive number whose value is determined in the following steps. Let $R(W)=\min\left\{I\left(W;Y_1,\hat{\mathbf{Y}}_2\right),I\left(W;\hat{\mathbf{Y}}_1,Y_2\right)\right\}$. Let $S_{[W]\delta}^{(n)}$ denote the set of all $\mathbf{w}\in\mathcal{W}^n$ sequences such that $\mathbf{w}\in A_\delta^{*(n)}(W)$ and $A_\delta^{*(n)}(U,V|\mathbf{w})$ is non-empty, as defined in [23, corollary 5.11]. From [23, corollary 5.11] we have that $||S_{[W]\delta}^{(n)}||\ge 2^{n(H(W)-\phi)}$, where $\phi\to 0$ as $\delta\to 0$ and $n\to\infty$.
- Pick $2^{n(R(W)-\epsilon)}$ sequences from $S_{[W]\delta}^{(n)}$ in a uniform and independent manner according to

$$\Pr(\mathbf{w})=\begin{cases}\frac{1}{||S_{[W]\delta}^{(n)}||} & ,\mathbf{w}\in S_{[W]\delta}^{(n)}\\ 0 & ,\text{otherwise.}\end{cases}$$

  Label these sequences with $l\in\mathcal{M}_0\triangleq\left\{1,2,...,2^{n(R(W)-\epsilon)}\right\}$.
- For each sequence $\mathbf{w}(l)$, $l\in\mathcal{M}_0$, consider the set $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$,$\delta'=\delta\max\{||\mathcal{U}||,||\mathcal{V}||\}$. Since the sequences $\mathbf{w}\in\mathcal{W}^n$ are selected such that $A_\delta^{*(n)}(U,V|\mathbf{w}(l))$ is non-empty and since $(\mathbf{u},\mathbf{v})\in A_\delta^{*(n)}(U,V|\mathbf{w}(l))$ implies $\mathbf{u}\in A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$, then also $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$ in non-empty, and by [23, theorem 5.9], $||A_{\delta'}^{*(n)}(U|\mathbf{w}(l))||\ge 2^{n(H(U|W)-\psi)}$, $\psi\to 0$ as $\delta'\to 0$ and $n\to\infty$.
- For each $l\in\mathcal{M}_0$ pick $2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{*(n)}(U|\mathbf{w}(l))$

according to

$$\Pr(\mathbf{u}|l)=\begin{cases}\frac{1}{||A_{\delta'}^{*(n)}(U|\mathbf{w}(l))||} & ,\mathbf{u}\in A_{\delta'}^{*(n)}(U|\mathbf{w}(l))\\ 0 & ,\text{otherwise.}\end{cases}$$

Label these sequences with $\mathbf{u}(i|l)$, $i\in\mathcal{Z}_1\triangleq\left\{1,2,...,2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}\right\}$. Similarly, pick $2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}$ sequences in a uniform and independent manner from $A_{\delta'}^{*(n)}(V|\mathbf{w}(l))$ according to

$$\Pr(\mathbf{v}|l)=\begin{cases}\frac{1}{||A_{\delta'}^{*(n)}(V|\mathbf{w}(l))||} & ,\mathbf{v}\in A_{\delta'}^{*(n)}(V|\mathbf{w}(l))\\ 0 & ,\text{otherwise.}\end{cases}$$

Label these sequences with $\mathbf{v}(j|l)$, $j\in\mathcal{Z}_2\triangleq\left\{1,2,...,2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}\right\}$. δ is selected such that $||S_{[W]\delta}^{(n)}||\geq 2^{n(R(W)-\epsilon)}$, and $\forall l\in\mathcal{M}_0$ we have that $||A_{\delta'}^{*(n)}(U|\mathbf{w}(l))||\geq 2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-\epsilon)}$ and $||A_{\delta'}^{*(n)}(V|\mathbf{w}(l))||\geq 2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-\epsilon)}$.

- Partition the set $\mathcal{Z}_1$ into 2^{nR_1} subsets B_{w_1}, $w_1\in\mathcal{M}_1=\left\{1,2,...,2^{nR_1}\right\}$, let $B_{w_1}=\left[(w_1-1)2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-R_1-\epsilon)}+1,w_1 2^{n(I(U;Y_1,\hat{\mathbf{Y}}_2|W)-R_1-\epsilon)}\right]$. Similarly partition the set $\mathcal{Z}_2$ into 2^{nR_2} subsets C_{w_2}, $w_2\in\mathcal{M}_2=\left\{1,2,...,2^{nR_2}\right\}$, let $C_{w_2}=\left[(w_2-1)2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-R_2-\epsilon)}+1,w_2 2^{n(I(V;\hat{\mathbf{Y}}_1,Y_2|W)-R_2-\epsilon)}\right]$.
- For each triplet (l,w_1,w_2) consider the set

$$\mathcal{D}(w_1,w_2|l)\triangleq\left\{(m_1,m_2):m_1\in B_{w_1},m_2\in C_{w_2},(\mathbf{u}(m_1|l),\mathbf{v}(m_2|l))\in A_{\delta'}^{*(n)}(U,V|\mathbf{w}(l))\right\}.$$

By [29, lemma on pg. 121], we have that taking n large enough we can make $\Pr\left(||\mathcal{D}(w_1,w_2|l)||=0\right)\leq\epsilon$ for any arbitrary $\epsilon>0$, as long as

$$R_1\leq I(U;Y_1,\hat{\mathbf{Y}}_2|W) \tag{48a}$$

$$R_2\leq I(V;\hat{\mathbf{Y}}_1,Y_2|W) \tag{48b}$$

$$R_1+R_2\leq I(U;Y_1,\hat{\mathbf{Y}}_2|W)+I(V;\hat{\mathbf{Y}}_1,Y_2|W)-I(U;V|W). \tag{48c}$$

Note that the individual rate constraints are required to guarantee that the sets B_{w_1} and C_{w_2} are non-empty.
- For each $l\in\mathcal{M}_0$, we pick a unique pair of $(m_1(w_1,w_2,l),m_2(w_1,w_2,l))\in\mathcal{D}(w_1,w_2|l)$, $(w_1,w_2)\in\mathcal{M}_1\times\mathcal{M}_2$. The transmitter generates the codeword $\mathbf{x}(l,w_1,w_2)$ according to $p(\mathbf{x}(l,w_1,w_2))=\prod_{i=1}^n p(x_i|u_i(m_1(w_1,w_2,l)),v_i(m_2(w_1,w_2,l)),w_i(l))$. When transmitting the triplet (l,w_1,w_2) the transmitter outputs $\mathbf{x}(l,w_1,w_2)$.

3) Codebook Generation at the Receivers:

- For the first conference step from R_{x1} to R_{x2}, R_{x1} generates a codebook with $2^{nR_{12}^{\prime(1)}}$ codewords indexed by $z_{12}^{(1)}\in\mathcal{Z}_{12}^{(1)}=\left\{1,2,...,2^{nR_{12}^{\prime(1)}}\right\}$ according to the distribution $p\left(\hat{y}_1^{(1)}\right)$: $p\left(\hat{\mathbf{y}}_1^{(1)}(z_{12}^{(1)})\right)=\prod_{i=1}^n p\left(\hat{y}_{1,i}^{(1)}(z_{12}^{(1)})\right)$. R_{x1} uniformly and independently partitions the message set $\mathcal{Z}_{12}^{(1)}$ into $2^{nR_{12}^{(1)}}$ subsets indexed by $w_{12}^{(1)}\in\mathcal{W}_{12}^{(1)}=\left\{1,2,...,2^{nR_{12}^{(1)}}\right\}$. Denote these subsets with $\mathcal{S}_{12,w_{12}^{(1)}}^{(1)}$.
- For the first conference step from R_{x2} to R_{x1}, R_{x2} generates a codebook with $2^{nR_{21}^{\prime(1)}}$ codewords indexed by $z_{21}^{(1)}\in\mathcal{Z}_{21}^{(1)}=\left\{1,2,...,2^{nR_{21}^{\prime(1)}}\right\}$ for each codeword $\hat{\mathbf{y}}_1^{(1)}(z_{12}^{(1)})$, $z_{12}^{(1)}\in\mathcal{Z}_{12}^{(1)}$, in an i.i.d. manner according

to $p\left(\hat{\mathbf{y}}_2^{(1)}(z_{21}^{(1)}|z_{12}^{(1)})\right) = \prod_{i=1}^n p\left(\hat{y}_{2,i}^{(1)}(z_{21}^{(1)}|z_{12}^{(1)})\middle|\hat{y}_{1,i}^{(1)}(z_{12}^{(1)})\right)$. R_{x2} uniformly and independently partitions the message set $\mathcal{Z}_{21}^{(1)}$ into $2^{nR_{21}^{(1)}}$ subsets indexed by $w_{21}^{(1)} \in \mathcal{W}_{21}^{(1)} = \left\{1, 2, ..., 2^{nR_{21}^{(1)}}\right\}$. Denote these subsets with $\mathcal{S}^{(1)}_{21,w_{21}^{(1)}}$.

- For the k'th conference step from R_{x1} to R_{x2}, R_{x1} considers each combination of $z_{12}^{(1)}, z_{12}^{(2)}, ..., z_{12}^{(k-1)}$, $z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-1)}$. For each combination, R_{x1} generates a codebook with $2^{nR_{12}'^{(k)}}$ messages indexed by $z_{12}^{(k)} \in \mathcal{Z}_{12}^{(k)} = \left\{1, 2, ..., 2^{nR_{12}'^{(k)}}\right\}$, according to the distribution $p\left(\hat{y}_1^{(k)}|\hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(k-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(k-1)}\right)$. R_{x1} uniformly and independently partitions the message set $\mathcal{Z}_{12}^{(k)}$ into $2^{nR_{12}^{(k)}}$ subsets indexed by $w_{12}^{(k)} \in \mathcal{W}_{12}^{(k)} = \left\{1, 2, ..., 2^{nR_{12}^{(k)}}\right\}$. Denote these subsets with $\mathcal{S}^{(k)}_{12,w_{12}^{(k)}}$.
- The codebook for the k'th conference step from R_{x2} to R_{x1} is generated in a parallel manner for each combination of $z_{12}^{(1)}, z_{12}^{(2)}, ..., z_{12}^{(k)}$, $z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-1)}$.

4) Decoding and Encoding at R_{x1} at the k'th Conference Cycle ($k \le K$) for Transmission Block i: R_{x1} needs first to decode the message $z_{21}^{(k-1)}$ sent from R_{x2} at the $(k-1)$'th cycle. To that end, R_{x1} uses $w_{21}^{(k-1)}$, the index received from R_{x2} at the $(k-1)$'th conference step. In decoding $z_{21}^{(k-1)}$ we assume that all the previous $z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-2)}$ were correctly decoded at R_{x1}. We denote the $\hat{\mathbf{y}}_2^{(k)}$ sequences corresponding to $z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-2)}$ by $\hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), ..., \hat{\mathbf{y}}_2(k-2)$, and similarly define $\hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), ..., \hat{\mathbf{y}}_1(k-1)$.

- R_{x1} first generates the set $\mathcal{L}_1(k-1)$ defined by:

$$\mathcal{L}_1(k-1) = \Big\{ z_{21}^{(k-1)} \in \mathcal{Z}_{21}^{(k-1)} : \Big(\hat{\mathbf{y}}_2^{(k-1)}(z_{21}^{(k-1)}|z_{12}^{(1)}, z_{12}^{(2)}, ..., z_{12}^{(k-1)}, z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-2)}),$$

$$\hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), ..., \hat{\mathbf{y}}_1(k-1), \hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), ..., \hat{\mathbf{y}}_2(k-2), \mathbf{y}_1(i) \Big) \in A_\epsilon^{*(n)} \Big\}.$$

- R_{x1} then looks for a unique $z_{21}^{(k-1)} \in \mathcal{Z}_{21}^{(k-1)}$ such that $z_{21}^{(k-1)} \in \mathcal{L}_1(k-1) \bigcap \mathcal{S}^{(k-1)}_{21,w_{21}^{(k-1)}}$. If there is none or there is more than one, an error is declared.
- From an argument similar to [30], the probability of error can be made arbitrarily small by taking n large enough as long as

$$R_{21}'^{(k-1)} < I\left(\hat{Y}_2^{(k-1)}; Y_1 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-2)}\right) + R_{21}^{(k-1)} - \epsilon.$$

Here, $k > 1$, since for the first conference message from R_{x1} to R_{x2} no decoding takes place.

In generating the k'th conference message to R_{x2}, it is assumed that all the previous $k-1$ messages from R_{x2} were decoded correctly.

- R_{x1} looks for a message $z_{12}^{(k)} \in \mathcal{Z}_{12}^{(k)}$ such that

$$\Big(\hat{\mathbf{y}}_1^{(k)}(z_{12}^{(k)}|z_{12}^{(1)}, z_{12}^{(2)}, ..., z_{12}^{(k-1)}, z_{21}^{(1)}, z_{21}^{(2)}, ..., z_{21}^{(k-1)}),$$

$$\hat{\mathbf{y}}_1(1), \hat{\mathbf{y}}_1(2), ..., \hat{\mathbf{y}}_1(k-1), \hat{\mathbf{y}}_2(1), \hat{\mathbf{y}}_2(2), ..., \hat{\mathbf{y}}_2(k-1), \mathbf{y}_1(i) \Big) \in A_\epsilon^{*(n)}.$$

From the argument in [30], the probability that such a sequence exists can be made arbitrarily close to 1 by taking n large enough as long as

$$R_{12}'^{(k)} > I\left(\hat{Y}_1^{(k)}; Y_1 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + \epsilon.$$

- R_{x1} looks for the partition of $\mathcal{Z}_{12}^{(k)}$ into which $z_{12}^{(k)}$ belongs. Denote the index of this partition with $w_{12}^{(k)}$.
- R_{x1} transmits $w_{12}^{(k)}$ to R_{x2} through the conference link.

5) Decoding and Encoding at R_{x2} at the k'th Conference Step ($k \leq K$) for Transmission Block i: Using similar arguments to section V-B.4, we obtain the following rate constraints:

- Decoding $z_{12}^{(k)}$ at R_{x2} can be done with an arbitrarily small probability of error by taking n large enough as long as
$$R_{12}'^{(k)} < I\left(\hat{Y}_1^{(k)}; Y_2 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + R_{12}^{(k)} - \epsilon.$$
- Encoding $z_{21}^{(k)}$ can be done with an arbitrarily small probability of error by taking n large enough as long as
$$R_{21}'^{(k)} > I\left(\hat{Y}_2^{(k)}; Y_2 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + \epsilon.$$

6) Combining All Conference Rate Bounds: First consider the bounds on $R_{12}'^{(k)}$, $k = 1, 2, ..., K$:

$$\begin{aligned}I\left(\hat{Y}_1^{(k)}; Y_1 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + \epsilon < R_{12}'^{(k)} < \\ I\left(\hat{Y}_1^{(k)}; Y_2 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + R_{12}^{(k)} - \epsilon.\end{aligned}$$

This can be satisfied only if

$$\begin{aligned}&I\left(\hat{Y}_1^{(k)}; Y_2 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + R_{12}^{(k)} - \epsilon > \\ &\qquad I\left(\hat{Y}_1^{(k)}; Y_1 \middle| \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + \epsilon \\ \Rightarrow\ & R_{12}^{(k)} > H\left(\hat{Y}_1^{(k)} \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) \\ &\qquad - H\left(\hat{Y}_1^{(k)} \middle| Y_1, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + 2\epsilon \\ &\qquad = I\left(\hat{Y}_1^{(k)}; Y_1 \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + 2\epsilon.\end{aligned}$$

Hence

$$\begin{aligned}C_{12} &= \sum_{k=1}^{K} R_{12}^{(k)} \\ &\geq \sum_{k=1}^{K} \left(I\left(\hat{Y}_1^{(k)}; Y_1 \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + 2\epsilon \right) \\ &= \sum_{k=1}^{K} \Bigg[I\left(\hat{Y}_1^{(k)}; Y_1 \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) \\ &\qquad + I\left(\hat{Y}_2^{(k)}; Y_1 \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) \Bigg] + 2K\epsilon \\ &= \sum_{k=1}^{K} I\left(\hat{Y}_1^{(k)}, \hat{Y}_2^{(k)}; Y_1 \middle| Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(k-1)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(k-1)}\right) + 2K\epsilon \\ &= I\left(\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K)}; Y_1 \middle| Y_2\right) + 2K\epsilon, \end{aligned} \tag{49}$$

and similarly

$$C_{21} \geq I\left(\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K)}; Y_2 \middle| Y_1\right) + 2K\epsilon. \tag{50}$$

This provides the rate constraints on the conference auxiliary variables of (46a) and (46b).

7) Decoding at R_{x1}: R_{x1} uses $\mathbf{y}_1(i)$ and $\hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, ..., \hat{\mathbf{y}}_2^{(K)}$ received from R_{x2}, to decode $(l_i, w_{1,i})$ as follows:

- R_{x1} looks for a unique message $l \in \mathcal{M}_0$ such

$$\left(\mathbf{w}(l), \mathbf{y}_1(i), \hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, ..., \hat{\mathbf{y}}_2^{(K)}\right) \in A_\epsilon^{*(n)}.$$

From the point-to-point channel capacity theorem (see [29]), this can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_0 \leq I(W; Y_1, \hat{\mathbf{Y}}_2). \tag{51}$$

Denote the decoded message $\hat{l}_i$. Now R_{x1} decodes $w_{1,i}$ by looking for a unique $k \in \mathcal{Z}_1$ such that

$$\left(\mathbf{u}(k|\hat{l}_i), \mathbf{w}(\hat{l}_i), \mathbf{y}_1(i), \hat{\mathbf{y}}_2^{(1)}, \hat{\mathbf{y}}_2^{(2)}, ..., \hat{\mathbf{y}}_2^{(K)}\right) \in A_\epsilon^{*(n)}.$$

If a unique such k exists, then denote the decoded index with $\hat{k} = k$. Now R_{x1} looks for the partition of $\mathcal{Z}_1$ into which $\hat{k}$ belongs and sets $\hat{w}_{1,i}$ to be the index of that partition: $\hat{k} \in B_{\hat{w}_{1,i}}$. Similarly to the proof in [24, ch 14.6.2], assuming successful decoding of l_i, the probability of error can be made arbitrarily small by taking n large enough as long as

$$\frac{1}{n} \log_2 ||\mathcal{Z}_1|| \leq I(U; Y_1, \hat{\mathbf{Y}}_2 | W),$$

which is satisfied by construction.

8) Decoding at R_{x2}: Repeating similar steps for decoding at R_{x2} we get that decoding l_i can be done with an arbitrarily small probability of error by taking n large enough as long as

$$R_0 \leq I(W; \hat{\mathbf{Y}}_1, Y_2), \tag{52}$$

and assuming successful decoding of l_i, decoding $w_{2,i}$ with an arbitrarily small probability of error requires that

$$\frac{1}{n} \log_2 ||\mathcal{Z}_2|| \leq I(V; \hat{\mathbf{Y}}_1, Y_2 | W),$$

which again is satisfied by construction.

Finally, collecting (48a), (48b), (48c), (51) and (52) give the achievable rate constraints of theorem 4, and (49) and (50) give the conference rate constraints of the theorem. ■

C. The Cooperative Broadcast Channel with a Single Common Message

In the single common message cooperative broadcast scenario, a single transmitter sends a message to two receivers encoded in a single channel codeword X^n. This scenario is depicted in figure 12. After conferencing, each receiver decodes the message. For this setup we have the following upper bound:

Proposition 3: ([27, theorem 6]) Consider the general broadcast channel $(\mathcal{X}, p(y_1, y_2|x), \mathcal{Y}_1 \times \mathcal{Y}_2)$ with cooperating receivers having noiseless conference links of finite capacities C_{12} and C_{21} between them. Then, for sending a common message to both receivers, any rate R must satisfy

$$R \leq \sup_{p_X(x)} \min\left\{ I(X; Y_1) + C_{21}, I(X; Y_2) + C_{12}, I(X; Y_1, Y_2) \right\}.$$

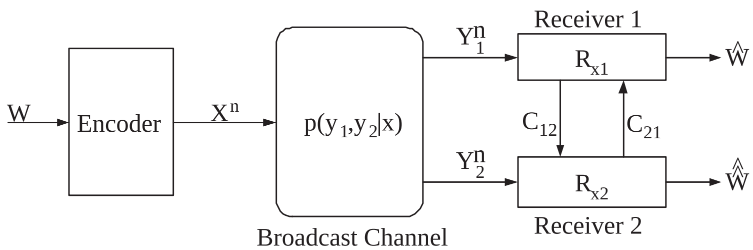

Fig. 12. The broadcast channel with cooperating receivers, for the single common message case. $\hat{W}$ and $\hat{\hat{W}}$ are the estimates of W at R_{x1} and R_{x2} respectively.

In [27] we also derived the following achievable rate for this scenario:

Proposition 4: ([27, theorem 5]) Assume the broadcast channel setup of proposition 3. Then, for sending a common message to both receivers, any rate R satisfying

$$R \leq \sup_{p_X(x)} \Big[\max \Big\{ R_{12}(p_X(x)), R_{21}(p_X(x)) \Big\} \Big],$$

$$R_{12}(p_X(x)) \triangleq \min \Big(I(X;Y_1) + C_{21}, \max \big\{ I(X;Y_2), I(X;Y_2) - H(Y_1|Y_2,X) + \min \big(C_{12}, H(Y_1|Y_2) \big) \big\} \Big), \quad \text{(53a)}$$

$$R_{21}(p_X(x)) \triangleq \min \Big(I(X;Y_2) + C_{12}, \max \big\{ I(X;Y_1), I(X;Y_1) - H(Y_2|Y_1,X) + \min \big(C_{21}, H(Y_2|Y_1) \big) \big\} \Big), \quad \text{(53b)}$$

is achievable.

Note that this rate expression depends only on the parameters of the problem and is, therefore, computable. In proposition 4 the achievable rate increases linearly with the cooperation capacity. The downside of this method is that it produces a rate increase over the non-cooperative rate only for conference links capacities that exceed some minimum values.

Specializing the three independent messages result to the single common message case we obtain the following achievable rate with a K-cycle conference for the general BC with a single common message:

Corollary 3: Consider the general broadcast channel with cooperating receivers, having noiseless conference links of finite capacities C_{12} and C_{21} between them. Let the receivers hold a conference that consists of K cycles. Then, any rate R satisfying

$$R = \max \{ R_{12}, R_{21} \}, \quad (54)$$

is achievable.

Here R_{12} is defined as follows:

$$R_{12} = \sup_{p_X(x), \alpha \in [0,1]} \min \{ R_1, R_2 \}, \quad (55)$$

with

$$R_1 = I\left(X; Y_1, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \right) + \alpha C_{21}, \quad \text{(56a)}$$

$$R_2 = I\left(X; Y_2, \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)} \right), \quad \text{(56b)}$$

subject to

$$C_{12} \geq I\left(Y_1; \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \middle| Y_2\right), \tag{57a}$$

$$(1-\alpha)C_{21} \geq I\left(Y_2; \hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)} \middle| Y_1\right), \tag{57b}$$

for the joint distribution

$$\begin{aligned} &p\left(x, y_1, y_2, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(K)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(K-1)}\right) = \\ &\quad p(x)p(y_1, y_2|x)p\left(\hat{y}_1^{(1)}|y_1\right) p\left(\hat{y}_2^{(1)}|y_2, \hat{y}_1^{(1)}\right) \cdots p\left(\hat{y}_1^{(k)}|y_1, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(k-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(k-1)}\right) \times \\ &\quad p\left(\hat{y}_2^{(k)}|y_2, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(k)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(k-1)}\right) \cdots p\left(\hat{y}_2^{(K-1)}|y_2, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(K-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(K-2)}\right) \\ &\quad \times p\left(\hat{y}_1^{(K)}|y_1, \hat{y}_1^{(1)}, \hat{y}_1^{(2)}, ..., \hat{y}_1^{(K-1)}, \hat{y}_2^{(1)}, \hat{y}_2^{(2)}, ..., \hat{y}_2^{(K-1)}\right). \end{aligned}$$

The cardinality of the k'th auxiliary random variables are bounded by:

$$\begin{aligned} ||\hat{\mathcal{Y}}_1^{(k)}|| &\leq ||\mathcal{Y}_1|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_2^{(l)}|| + 1, \qquad k = 1, 2, ..., K \\ ||\hat{\mathcal{Y}}_2^{(k)}|| &\leq ||\mathcal{Y}_2|| \times \prod_{l=1}^{k} ||\hat{\mathcal{Y}}_1^{(l)}|| \times \prod_{l=1}^{k-1} ||\hat{\mathcal{Y}}_2^{(l)}|| + 1, \qquad k = 1, 2, ..., K-1. \end{aligned}$$

R_{21} is defined in a parallel manner to R_{12}, with R_{x2} performing the first conference step, and the appropriate change in the probability chain.

The proof of corollary 3 is provided in appendix B.

We note that [12, theorem 2] presents a similar result for this scenario, under the constraint that the memoryless broadcast channel can be decomposed as $p(\mathbf{y}_1, \mathbf{y}_2|\mathbf{x}) = \prod_{i=1}^n p(y_{1,i}|x_i)p(y_{2,i}|x_i)$, and considering the sum-rate of the conference. Here we show that the same achievable rate expressions hold for the general memoryless broadcast channel. A recent result appears in [31], where the single common message case for a Gaussian BC is considered. In the multi-cycle conference considered in this section, we let the auxiliary RVs follow a more general chain than that of [31] — which results in a larger achievable rate.

D. A Single-Cycle Conference with TS-EAF

Consider the case where only a single cycle of conferencing between the receivers is allowed. Specializing corollary 3 to a single cycle case we obtain

$$R_1 = I(X; Y_1) + C_{21} \tag{58a}$$

$$R_2 = I(X; Y_2, \hat{Y}_1^{(1)}) \tag{58b}$$

$$C_{12} \geq I(Y_1; \hat{Y}_1^{(1)}|Y_2), \tag{58c}$$

and the TS-EAF assignment is

$$p(\hat{y}_1^{(1)}|y_1) = \begin{cases} q_1, & \hat{y}_1^{(1)} = y_1 \\ 1 - q_1, & \hat{y}_1^{(1)} = \Omega \notin \mathcal{Y}_1. \end{cases}$$

Applying the TS-EAF assignment to (58c) and (58b) we obtain

$$\begin{aligned}
C_{12} &\geq I(Y_1;\hat{Y}_1^{(1)}|Y_2)\\
&= H(Y_1|Y_2) - H(Y_1|Y_2,\hat{Y}_1^{(1)})\\
&= H(Y_1|Y_2) - q_1 H(Y_1|Y_2,Y_1) - (1-q_1)H(Y_1|Y_2)\\
&= q_1 H(Y_1|Y_2)\\
R_2 &= I(X;Y_2,\hat{Y}_1^{(1)})\\
&= I(X;Y_2) + H(X|Y_2) - H(X|Y_2,\hat{Y}_1^{(1)})\\
&= I(X;Y_2) + H(X|Y_2) - (1-q_1)H(X|Y_2) - q_1 H(X|Y_2,Y_1)\\
&= I(X;Y_2) + q_1 I(X;Y_1|Y_2).
\end{aligned}$$

Maximizing R_2 requires maximizing $q_1 \in [0,1]$. Therefore setting $q_1 = \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^*$, we obtain $R_2 = I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2)$. Combining with R_1 we have that the rate when R_{x2} decodes first is given by

$$R_{12} = \min\left\{ I(X;Y_1) + C_{21}, I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2) \right\},$$

and by symmetric argument we can obtain R_{21}. We conclude that the rate for the single-cycle conference with TS-EAF is given by

$$\begin{aligned}
R &= \sup_{p(x)} \min\left\{R_{12}, R_{21}\right\},\\
R_{12} &= \min\left\{ I(X;Y_1) + C_{21}, I(X;Y_2) + \left[\frac{C_{12}}{H(Y_1|Y_2)}\right]^* I(X;Y_1|Y_2) \right\}\\
R_{21} &= \min\left\{ I(X;Y_1) + \left[\frac{C_{21}}{H(Y_2|Y_1)}\right]^* I(X;Y_2|Y_1), I(X;Y_2) + C_{12} \right\}.
\end{aligned}$$

We note that this rate is always better than the point-to-point rate and also better than the joint-decoding rate of proposition 4 (whenever cooperation can provide a rate increase). However, as in proposition 4, at least one receiver has to satisfy the Slepian-Wolf condition for the full cooperation rate to be achieved. We also note that using TS-EAF with more than two steps does not improve upon this result.

Finally, we demonstrate the results of proposition 4 and corollary 3 through a symmetric BC example: consider the symmetric broadcast channel where $\mathcal{Y}_1 = \mathcal{Y}_2 = \mathcal{Y}$ and

$$p_{Y_1|Y_2,X}(a|b,x) = p_{Y_2|Y_1,X}(a|b,x),$$

for any $a,b \in \mathcal{Y}\times\mathcal{Y}$ and $x \in \mathcal{X}$. Let $C_{21} = C_{12} = C$. For this scenario we have that $R_{12} = R_{21}$, in corollary 3 and also $R_{12}(p_X(x)) = R_{21}(p_X(x))$ in proposition 4. The resulting rate is depicted in figure 13 for a fixed probability $p(x)$. We can see that for this case, time-sharing exceeds joint-decoding for all values of C. Both methods meet the upper bound at $C = H(Y_1|Y_2)$. We note that this is a corrected version of the figure in [32].

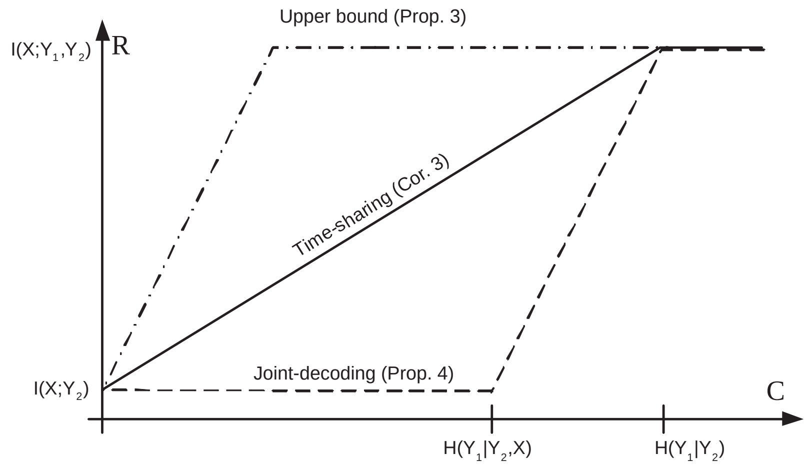

Fig. 13. The achievable rate R vs. conference capacity C, for proposition 3 (dashed-dot), proposition 4 (dashed) and corollary 3 (solid), for the symmetric broadcast channel.

VI. Conclusions

In this paper we considered the EAF technique using time-sharing on the auxiliary RVs. We first showed that incorporating joint-decoding at the destination into the EAF technique results in a special case of the classic EAF of [2, theorem 6]. We then used the time-sharing assignment of the auxiliary RVs to obtain an easily computable achievable rate for the multiple-relay case, which can be compared against the DAF-based results, to select the highest rate for any given scenario. Next, we showed that for the Gaussian relay channel with coded modulation, the Gaussian auxiliary RV assignment is not always optimal, and a TS-EAF implementing a per-symbol hard decision may sometimes perform better. Finally, we considered a third application of TS-EAF to the cooperative broadcast scenario with a multi-cycle conference. We first derived an achievable rate for the general channel, and then we specialized it to the single-cycle conference for which we obtained an explicit achievable rate. This rate is superior to the explicit expression that can be obtained with joint-decoding.

VII. Acknowledgements

In the final version.

Appendix A
Expressions for Section IV

A. Hard-Decision Estimate-and-Forward

We evaluate $I(X;\hat{Y}_1,Y)$, with $p(\hat{Y}_1|Y_1)$ given by (40a) and (40b) using:

$$I(X;\hat{Y}_1,Y) = I(X;\hat{Y}_1) + I(X;Y|\hat{Y}_1).$$

1) Evaluating $I(X;\hat{Y}_1)$: Note that both X and $\hat{Y}_1$ are discrete RVs, therefore $I(X;\hat{Y}_1)$ can be evaluated using the discrete entropies. The conditional distribution of $\hat{Y}_1$ given X is given by:

$$p(\hat{Y}_1|X=\sqrt{P}) = \begin{cases} P_1 \cdot P_{\text{no erase}}, & 1 \\ 1 - P_{\text{no erase}}, & E \\ (1-P_1)P_{\text{no erase}}, & -1 \end{cases} \tag{A.1}$$

where

$$P_1 = \Pr(Y_1 > 0|X=\sqrt{P}).$$

$p(\hat{Y}_1|X=-\sqrt{P})$ can be obtained from $p(\hat{Y}_1|X=\sqrt{P})$ by switching 1 and -1 in (A.1).

2) Evaluating $I(X;Y|\hat{Y}_1)$: write first

$$I(X;Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|\hat{Y}_1, X),$$

and we note that

$$h(Y|\hat{Y}_1, X) = h(X+N|\hat{Y}_1, X) = h(N|\hat{Y}_1, X) = h(N) = \frac{1}{2}\log_2(2\pi e\sigma^2).$$

Using the chain rule we write

$$h(Y|\hat{Y}_1) = p(\hat{Y}_1=1)h(Y|\hat{Y}_1=1) + p(\hat{Y}_1=E)h(Y|\hat{Y}_1=E) + p(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1),$$

$p(\hat{Y}_1)$ can be obtained by combining (38) and (A.1) which results in

$$p(\hat{Y}_1) = \begin{cases} \frac{1}{2}P_{\text{no erase}}, & 1 \\ 1 - P_{\text{no erase}}, & E \\ \frac{1}{2}P_{\text{no erase}}, & -1 \end{cases}, \tag{A.2}$$

and we note that $h(Y|\hat{Y}_1=E) = h(Y)$, since erasure is equivalent to no prior information. Finally we note that by definition

$$\begin{aligned} h(Y) &= -\int_{y=-\infty}^{\infty} f(y)\log_2(f(y))dy, \\ f(Y) &= \Pr(X=\sqrt{P})f(Y|X=\sqrt{P}) + \Pr(X=-\sqrt{P})f(Y|X=-\sqrt{P}) \\ &= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2) + G_y(-\sqrt{P},\sigma^2)\right), \end{aligned} \tag{A.3}$$

where

$$G_x(a,b) = \frac{1}{\sqrt{2\pi b}}e^{-\frac{(x-a)^2}{2b}}. \tag{A.4}$$

Next, we have

$$h(Y|\hat{Y}_1=1) = -\int_{y=-\infty}^{\infty} f(y|\hat{y}_1=1)\log_2(f(y|\hat{y}_1=1))dy \tag{A.5}$$

$$\begin{aligned} f(Y|\hat{Y}_1=1) &= \frac{f(Y,\hat{Y}_1=1)}{\Pr(\hat{Y}_1=1)} \\ &= \frac{f(Y,Y_1>0)P_{\text{no erase}}}{\Pr(Y_1>0)P_{\text{no erase}}} \\ &= \frac{f(Y,Y_1>0)}{\Pr(Y_1>0)}, \qquad \text{(A.6)} \\ f(Y,Y_1>0) &= \Pr(X=\sqrt{P})f(Y,Y_1>0|X=\sqrt{P}) + \Pr(X=-\sqrt{P})f(Y,Y_1>0|X=-\sqrt{P}) \\ &= \frac{1}{2}\left(f(Y,Y_1>0|X=\sqrt{P}) + f(Y,Y_1>0|X=-\sqrt{P})\right). \qquad \text{(A.7)} \end{aligned}$$

Using

$$f_{Y,Y_1}(y,y_1|x) = \mathcal{N}\left(\begin{pmatrix} x \\ g\cdot x\end{pmatrix}, \begin{pmatrix}\sigma^2 & 0 \\ 0 & \sigma_1^2\end{pmatrix}\right) = G_y(x,\sigma^2)G_{y_1}(g\cdot x,\sigma_1^2),$$

we obtain

$$f(Y,Y_1>0|X) = \int_{y_1=0}^{\infty} f(y,y_1|x)dy_1 = G_y(x,\sigma^2)\int_{y_1=0}^{\infty} G_{y_1}(g\cdot x,\sigma_1^2)dy_1.$$

Next we need to evaluate $I(\hat{Y}_1;Y_1|Y) = h(Y_1|Y) - h(Y_1|Y,\hat{Y}_1)$:

1) $h(Y_1|Y) = h(Y,Y_1) - h(Y)$. Here

$$\begin{aligned} h(Y,Y_1) &= -\int_{y=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty} f(y,y_1)\log_2(f(y,y_1))dy\,dy_1, \\ f(Y,Y_1) &= \frac{1}{2}\left(f(Y,Y_1|X=\sqrt{P}) + f(Y,Y_1|X=-\sqrt{P})\right), \\ f(Y,Y_1|X) &= G_y(x,\sigma^2)G_{y_1}(g\cdot x,\sigma_1^2). \end{aligned}$$

2) By the definition of conditional entropy we have

$$h(Y_1|Y,\hat{Y}_1) = p(\hat{Y}_1=1)h(Y_1|Y,\hat{Y}_1=1) + p(\hat{Y}_1=E)h(Y_1|Y,\hat{Y}_1=E) + p(\hat{Y}_1=-1)h(Y_1|Y,\hat{Y}_1=-1),$$

where $h(Y_1|Y,\hat{Y}_1=E) = h(Y_1|Y)$, and for $\hat{Y}_1=1$, for example, we have

$$h(Y_1|Y,\hat{Y}_1=1) = -\int_{y=-\infty}^{\infty}\int_{y_1=-\infty}^{\infty} f(y,y_1|\hat{y}_1=1)\log_2(f(y_1|y,\hat{y}_1=1))dy\,dy_1.$$

Finally, we need to derive the distributions $f(y,y_1|\hat{y}_1=1)$ and $f(y_1|y,\hat{y}_1=1)$. Begin with

$$\begin{aligned} f_{Y,Y_1|\hat{Y}_1}(y,y_1|\hat{y}_1=1) &= \frac{f_{Y,Y_1,\hat{Y}_1}(y,y_1,\hat{y}_1=1)}{\Pr(\hat{y}_1=1)} \\ &= \frac{f_{Y,Y_1,\hat{Y}_1}(y,y_1,y_1>0)P_{\text{no erase}}}{\Pr(y_1>0)P_{\text{no erase}}} = f(y,y_1|y_1>0) = \begin{cases} \frac{f_{Y,Y_1}(y,y_1)}{\Pr(Y_1>0)}, & y_1>0 \\ 0, & y_1\le 0\end{cases} \end{aligned}$$

and due to the symmetry, $\Pr(Y_1>0) = \Pr(Y_1\le 0) = \frac{1}{2}$. We also have

$$\begin{aligned} f(Y_1|Y,\hat{Y}_1=1) &= \frac{f(Y_1,Y|\hat{Y}_1=1)}{f(Y|\hat{Y}_1=1)} = \frac{f(Y_1,Y|Y_1>0)}{f(Y|Y_1>0)} = \frac{\frac{f(Y_1,Y)}{\Pr(Y_1>0)}}{\frac{f(Y,Y_1>0)}{\Pr(Y_1>0)}} = \frac{f(Y_1,Y)}{f(Y,Y_1>0)}, \quad Y_1>0 \\ f(Y_1|Y,\hat{Y}_1=1) &= 0, \quad Y_1\le 0. \end{aligned}$$

B. Evaluation of the Rate with DHD

We evaluate the achievable rate using $I(X;Y,\hat{Y}_1) = I(X;\hat{Y}_1) + I(X;Y|\hat{Y}_1)$. The distribution of $\hat{Y}_1$ is given by:

$$\begin{aligned}
\Pr(\hat{Y}_1 = 1) = \Pr(Y_1 > T) &= \frac{1}{2}\left(\Pr(Y_1 > T|X = \sqrt{P}) + \Pr(Y_1 > T|X = -\sqrt{P})\right) \\
&= \frac{1}{2}\left(\int_{y_1 > T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 + \int_{y_1 > T} G_{y_1}(-g\sqrt{P}, \sigma_1^2)dy_1\right) \\
\Pr(\hat{Y}_1 = E) = \Pr(|Y_1| \le T) &= \frac{1}{2}\left(\Pr(|Y_1| \le T|X = \sqrt{P}) + \Pr(|Y_1| \le T|X = -\sqrt{P})\right) \\
&= \frac{1}{2}\left(\int_{y_1 = -T}^{T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 + \int_{y_1 = -T}^{T} G_{y_1}(-g\sqrt{P}, \sigma_1^2)dy_1\right),
\end{aligned}$$

and by symmetry, $\Pr(\hat{Y}_1 = 1) = \Pr(\hat{Y}_1 = -1)$ and $H(\hat{Y}_1|X = \sqrt{P}) = H(\hat{Y}_1|X = -\sqrt{P})$. Therefore, we need the conditional distribution $p(\hat{Y}_1|X = \sqrt{P})$:

$$\begin{aligned}
\Pr(\hat{Y}_1 = 1|X = \sqrt{P}) &= \Pr(Y_1 > T|X = \sqrt{P}) = \int_{y_1 > T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 \\
\Pr(\hat{Y}_1 = -1|X = \sqrt{P}) &= \Pr(Y_1 < -T|X = \sqrt{P}) = \int_{y_1 < -T} G_{y_1}(g\sqrt{P}, \sigma_1^2)dy_1 \\
\Pr(\hat{Y}_1 = E|X = \sqrt{P}) &= 1 - \Pr(\hat{Y}_1 = 1|X = \sqrt{P}) - \Pr(\hat{Y}_1 = -1|X = \sqrt{P}).
\end{aligned}$$

This allows us to evaluate $I(X;\hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$. For evaluating $I(X;Y|\hat{Y}_1)$ note that

$$h(Y|\hat{Y}_1, X) = h(X + N|\hat{Y}_1, X) = h(N|\hat{Y}_1, X) = h(N) = \frac{1}{2}\log_2(2\pi e\sigma^2),$$

and we need only to evaluate $h(Y|\hat{Y}_1)$: by definition

$$h(Y|\hat{Y}_1) = \Pr(\hat{Y}_1 = 1)h(Y|\hat{Y}_1 = 1) + \Pr(\hat{Y}_1 = E)h(Y|\hat{Y}_1 = E) + \Pr(\hat{Y}_1 = -1)h(Y|\hat{Y}_1 = -1),$$

and note that $h(Y|\hat{Y}_1 = E) = h(Y)$. Finally,

$$\begin{aligned}
h(Y|\hat{Y}_1 = 1) &= -\int_{y=-\infty}^{\infty} f(y|\hat{y}_1 = 1)\log_2(f(y|\hat{y}_1 = 1))dy \\
f_{Y|\hat{Y}_1}(y|\hat{y}_1 = 1) &= f(y|y_1 > T) = \frac{f(y, y_1 > T)}{\Pr(Y_1 > T)} \\
f_{Y,Y_1}(y, y_1 > T) &= \frac{1}{2}\left(f(y, y_1 > T|X = \sqrt{P}) + f(y, y_1 > T|X = -\sqrt{P})\right) \\
&= \frac{1}{2}\left(G_y(\sqrt{P}, \sigma^2)\Pr(Y_1 > T|X = \sqrt{P}) + G_y(-\sqrt{P}, \sigma^2)\Pr(Y_1 > T|X = -\sqrt{P})\right).
\end{aligned}$$

Evaluating $I(\hat{Y}_1; Y_1|Y)$ we have:

$$\begin{aligned}
I(\hat{Y}_1; Y_1|Y) &= H(\hat{Y}_1|Y) - H(\hat{Y}_1|Y, Y_1) \\
&\overset{(a)}{=} H(\hat{Y}_1|Y) \\
&= H(\hat{Y}_1) + h(Y|\hat{Y}_1) - h(Y),
\end{aligned}$$

where (a) is due to the deterministic mapping from Y_1 to $\hat{Y}_1$, and $h(Y)$ can be evaluated using (A.3).

1) DHD when $T \to 0$: As $T \to 0$ we have that $\Pr(\hat{Y}_1 = E) \to 0$ and $\hat{Y}_1$ converges in distribution to a Bernoulli RV with probability $\frac{1}{2}$. Therefore

$$\begin{aligned} f(Y, \hat{Y}_1 = 1) &= \frac{1}{2}\left(G_y(\sqrt{P}, \sigma^2)\Pr(Y_1 > T|X = \sqrt{P}) + G_y(-\sqrt{P}, \sigma^2)\Pr(Y_1 > T|X = -\sqrt{P})\right) \\ &\overset{T\to 0}{\approx} \frac{1}{2}\left(G_y(\sqrt{P}, \sigma^2)\Pr(Y_1 > 0|X = \sqrt{P}) + G_y(-\sqrt{P}, \sigma^2)\Pr(Y_1 > 0|X = -\sqrt{P})\right) \\ &= \frac{1}{2}\left(G_y(\sqrt{P}, \sigma^2)P_+ + G_y(-\sqrt{P}, \sigma^2)(1 - P_+)\right), \end{aligned}$$

where $P_+ = \Pr(Y_1 > 0|X = \sqrt{P})$. Now, letting $g \to 0$ we have that $P_+ \to \frac{1}{2}$ and therefore

$$\begin{aligned} f(Y|\hat{Y}_1 = 1) &\overset{g\to 0, T\to 0}{\longrightarrow} f(Y) \\ \Rightarrow h(Y|\hat{Y}_1 = 1) &\overset{g\to 0, T\to 0}{\longrightarrow} h(Y). \end{aligned}$$

We conclude that as $g \to 0, T \to 0$, then $h(Y|\hat{Y}_1) \to h(Y)$ and therefore the $I(Y_1; \hat{Y}_1|Y)$ becomes

$$I(Y_1; \hat{Y}_1|Y) = H(\hat{Y}_1) + h(Y|\hat{Y}_1) - h(Y) \overset{g\to 0, T\to 0}{\longrightarrow} 1$$

Using the continuity of $I(Y_1; \hat{Y}_1|Y)$ we conclude that for small values of g, as T decreases then $I(Y_1; \hat{Y}_1|Y)$ is bounded from below. This implies that for small g and small C the feasibility is obtained only for large T, which in turn implies low rate.

C. Evaluating the Information Rate with TS-DHD

1) Evaluating $I(X; Y, \hat{Y}_1)$: We first write

$$I(X; Y, \hat{Y}_1) = I(X; \hat{Y}_1) + I(X; Y|\hat{Y}_1).$$

Evaluating $I(X; \hat{Y}_1) = H(\hat{Y}_1) - H(\hat{Y}_1|X)$ requires the marginal of $\hat{Y}_1$. Using the mapping defined in (42)we find the marginal distribution of $\hat{Y}_1$:

$$\Pr(\hat{Y}_1) = \begin{cases} 1, & (1 - P_{\text{erase}})\Pr(Y_1 > T) \\ E, & \Pr(|Y_1| \le T) + P_{\text{erase}}\Pr(|Y_1| > T) \\ -1, & (1 - P_{\text{erase}})\Pr(Y_1 < -T) \end{cases},$$

where

$$\begin{aligned} \Pr(Y_1 > T) = \Pr(Y_1 < -T) &= \int_{y_1=T}^{\infty} \frac{1}{2}\left[G_{y_1}(\sqrt{P}, \sigma_1^2) + G_{y_1}(-\sqrt{P}, \sigma_1^2)\right] dy_1 \\ \Pr(|Y_1| < T) &= \int_{y_1=-T}^{T} \frac{1}{2}\left[G_{y_1}(\sqrt{P}, \sigma_1^2) + G_{y_1}(-\sqrt{P}, \sigma_1^2)\right] dy_1. \end{aligned}$$

Also, due to symmetry we have that $H(\hat{Y}_1|X = \sqrt{P}) = H(\hat{Y}_1|X = -\sqrt{P})$, and therefore we need only to find the conditional $\Pr(\hat{Y}_1|X = \sqrt{P})$:

$$\Pr(\hat{Y}_1|X = \sqrt{P}) = \begin{cases} 1, & (1 - P_{\text{erase}})\Pr(Y_1 > T|X = \sqrt{P}) \\ E, & \Pr(|Y_1| \le T|X = \sqrt{P}) + P_{\text{erase}}\Pr(|Y_1| > T|X = \sqrt{P}) \\ -1, & (1 - P_{\text{erase}})\Pr(Y_1 < -T|X = \sqrt{P}) \end{cases},$$

and we note that $f_{Y_1|X}(y_1|x=\sqrt{P}) = G_{y_1}(\sqrt{P},\sigma_1^2)$.

Next, we need to evaluate $I(X;Y|\hat{Y}_1) = h(Y|\hat{Y}_1) - h(Y|\hat{Y}_1,X)$. We first note that

$$h(Y|\hat{Y}_1,X) = h(X+N|X,\hat{Y}_1) = h(N|X,\hat{Y}_1) = h(N) = \frac{1}{2}\log_2(2\pi e\sigma_1^2).$$

Lastly, we have

$$h(Y|\hat{Y}_1) = \Pr(\hat{Y}_1=1)h(Y|\hat{Y}_1=1) + \Pr(\hat{Y}_1=E)h(Y|\hat{Y}_1=E) + \Pr(\hat{Y}_1=-1)h(Y|\hat{Y}_1=-1).$$

We note that $h(Y|\hat{Y}_1=E) = h(Y)$ and that $h(Y|\hat{Y}_1=1)$ and $h(Y|\hat{Y}_1=-1)$ are calculated exactly as in appendix A-B for the DHD case.

2) Evaluating $I(\hat{Y}_1;Y_1|Y)$: Begin by writing

$$\begin{aligned} I(\hat{Y}_1;Y_1|Y) &= h(\hat{Y}_1|Y_1) - h(\hat{Y}_1|Y_1,Y) \\ &= h(Y|\hat{Y}_1) + H(\hat{Y}_1) - h(Y) - h(\hat{Y}_1|Y_1) \end{aligned}$$

where we used the fact that given Y_1, $\hat{Y}_1$ is independent of Y. All the terms in the above expressions have been calculated in the previous subsection, except $h(\hat{Y}_1|Y_1)$:

$$\begin{aligned} h(\hat{Y}_1|Y_1) &= \Pr(\hat{Y}_1>T)h(\hat{Y}_1|Y_1>T) + \Pr(|Y_1|\le T)h(\hat{Y}_1||Y_1|\le T) + \Pr(Y_1<-T)h(\hat{Y}_1|Y_1<-T) \\ &= \Pr(\hat{Y}_1>T)H(P_{\text{erase}},1-P_{\text{erase}}) + \Pr(\hat{Y}_1<-T)H(P_{\text{erase}},1-P_{\text{erase}}) \\ &= (1-P(|Y_1|\le T)H(P_{\text{erase}},1-P_{\text{erase}}). \end{aligned}$$

D. Gaussian-Quantization Estimate-and-Forward

Here the relay uses the assignment of equation (36):

$$\hat{Y}_1 = Y_1 + N_Q, \qquad N_Q \sim \mathcal{N}(0,\sigma_Q^2).$$

We first evaluate

$$I(X;Y,\hat{Y}_1) = h(Y,\hat{Y}_1) - h(Y,\hat{Y}_1|X):$$

1)

$$\begin{aligned} h(Y,\hat{Y}_1) &= -\int_{y=-\infty}^{\infty}\int_{\hat{y}_1=-\infty}^{\infty} f_{Y,\hat{Y}_1}(y,\hat{y}_1)\log_2(f_{Y,\hat{Y}_1}(y,\hat{y}_1))dy\, d\hat{y}_1 \\ f_{Y,\hat{Y}_1}(y,\hat{y}_1) &= \frac{1}{2}\left(G_y(\sqrt{P},\sigma^2)G_{\hat{y}_1}(g\sqrt{P},\sigma_1^2+\sigma_Q^2) + G_y(-\sqrt{P},\sigma^2)G_{\hat{y}_1}(-g\sqrt{P},\sigma_1^2+\sigma_Q^2)\right). \end{aligned} \tag{A.8}$$

2) We also have

$$\begin{aligned} h(Y,\hat{Y}_1|X) &= h(X+N, gX+N_1+N_Q|X) \\ &= h(N, N_1+N_Q|X) \\ &= h(N) + h(N_1+N_Q) \\ &= \frac{1}{2}\log_2\left((2\pi e)^2\sigma^2(\sigma_1^2+\sigma_Q^2)\right). \end{aligned}$$

Lastly we need to evaluate

$$I(\hat{Y}_1;Y_1|Y) = h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y_1,Y) = h(\hat{Y}_1,Y) - h(Y) - h(\hat{Y}_1|Y_1,Y),$$

where

$$h(\hat{Y}_1|Y_1,Y) = h(Y_1+N_Q|Y_1,Y) = h(N_Q|Y_1,Y) = h(N_Q) = \frac{1}{2}\log_2(2\pi e\sigma_Q^2).$$

E. Approximation of HD-EAF for $\sigma^2 \to \infty$

Using (A.1) and (A.2) we can write

$$\begin{aligned}
R \leq I(X;\hat{Y}_1) &= H(\hat{Y}_1) - H(\hat{Y}_1|X) \\
&= H\left(\frac{1}{2}P_{\text{no erase}}, 1-P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right) - H\left(P_1 P_{\text{no erase}}, 1-P_{\text{no erase}}, (1-P_1)P_{\text{no erase}}\right) \\
&= -P_{\text{no erase}}\log_2\left(\frac{1}{2}P_{\text{no erase}}\right) - (1-P_{\text{no erase}})\log_2(1-P_{\text{no erase}}) + P_1 P_{\text{no erase}}\log_2(P_1 P_{\text{no erase}}) \\
&\quad +(1-P_{\text{no erase}})\log_2(1-P_{\text{no erase}}) + (1-P_1)P_{\text{no erase}}\log_2((1-P_1)P_{\text{no erase}}) \\
&= -P_{\text{no erase}}\log_2\left(P_{\text{no erase}}\right) + P_{\text{no erase}} + P_1 P_{\text{no erase}}\log_2(P_1) + P_1 P_{\text{no erase}}\log_2(P_{\text{no erase}}) \\
&\quad +(1-P_1)P_{\text{no erase}}\log_2(1-P_1) + (1-P_1)P_{\text{no erase}}\log_2(P_{\text{no erase}}) \\
&= P_{\text{no erase}}(1 + P_1\log_2(P_1) + (1-P_1)\log_2(1-P_1)) \\
&= P_{\text{no erase}}(1 - H(P_1, 1-P_1)).
\end{aligned}$$

$$\begin{aligned}
I(Y_1;\hat{Y}_1|Y) &= h(\hat{Y}_1|Y) - h(\hat{Y}_1|Y_1,Y) \\
&\overset{(a)}{\approx} H(\hat{Y}_1) - H(\hat{Y}_1|Y_1) \\
&= H\left(\frac{1}{2}P_{\text{no erase}}, 1-P_{\text{no erase}}, \frac{1}{2}P_{\text{no erase}}\right) - H(P_{\text{no erase}}, 1-P_{\text{no erase}}) \\
&= -2\frac{1}{2}P_{\text{no erase}}\log_2\left(\frac{1}{2}P_{\text{no erase}}\right) - (1-P_{\text{no erase}})\log_2\left(1-P_{\text{no erase}}\right) + P_{\text{no erase}}\log_2(P_{\text{no erase}}) \\
&\quad +(1-P_{\text{no erase}})\log_2\left(1-P_{\text{no erase}}\right) \\
&= P_{\text{no erase}},
\end{aligned}$$

where in (a) we used the fact that $\hat{Y}_1$ and Y are independent as $\sigma^2 \to \infty$, and that given Y_1, $\hat{Y}_1$ is independent of Y.

APPENDIX B
PROOF OF COROLLARY 3

In the following we highlight only the modifications from the general broadcast result due to the application of DAF to the last conference step from R_{x1} to R_{x2}, and the fact that we transmit a single message.

1) Codebook Generation and Encoding at the Transmitter: The transmitter generates 2^{nR} codewords $\mathbf{x}$ in an i.i.d. manner according to $p(\mathbf{x}(w)) = \prod_{i=1}^n p(x_i(w))$, $w \in \mathcal{W} = \left\{1, 2, ..., 2^{nR}\right\}$. For transmission of the message w_i at time i the transmitter outputs $\mathbf{x}(w_i)$.

2) Codebook Generation at the R_{x1}: The K conference steps from R_{x1} to R_{x2} are carried out exactly as in section V-B.4. The first $K-1$ steps from R_{x2} to R_{x1} are carried out as in section V-B.5. The K'th conference step from R_{x2} to R_{x1}, is different from that of theorem 4, as after the K'th step from R_{x1} to R_{x2}, R_{x2} may decode the message since R_{x2} received all the K conference messages from R_{x1}. Then, R_{x2} uses decode-and-forward for its K'th conference transmission to R_{x1}. Therefore, R_{x2} simply partitions $\mathcal{W}$ into $2^{n\alpha C_{21}}$ subsets in a uniform and independent manner.

3) Encoding and Decoding at the K'th Conference Step from R_{x2} to R_{x1}:

- Before the K'th conference step, R_{x2} decodes its message using his channel input and all the K conference messages received from R_{x1}. This can be done with an arbitrarily small probability of error as long as (56b) is satisfied.
- Having decoded its message, R_{x2} uses the decode-and-forward strategy to select the K'th conference message to R_{x1}. The conference capacity allocated to this step is $R_{21}^{(K)} = \alpha C_{21}$.
- Having received the K'th conference message from R_{x2}, R_{x1} can now decode its message using the information received at the first $K-1$ steps, and combining it with the information from the last step using the decode-and-forward decoding rule. This gives rise to (56a).

4) Combining All the Conference Rate Bounds: The bounds on $R_{12}'^{(k)}$, $k = 1, 2, ..., K$ can be obtained as in section V-B.6:

$$\begin{aligned} C_{12} &= \sum_{k=1}^{K} R_{12}^{(k)} \\ &\geq I\left(\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)}; Y_1 \middle| Y_2\right) + 2K\epsilon, \end{aligned}$$

and similarly

$$(1-\alpha)C_{21} \geq I\left(\hat{Y}_1^{(1)}, \hat{Y}_1^{(2)}, ..., \hat{Y}_1^{(K)}, \hat{Y}_2^{(1)}, \hat{Y}_2^{(2)}, ..., \hat{Y}_2^{(K-1)}; Y_2 \middle| Y_1\right) + 2K\epsilon,$$

where $(1-\alpha)C_{21}$ is the total capacity allocated to the first $K-1$ conference steps from R_{x2} to R_{x1}. This provides the rate constraints on the conference auxiliary variables.

REFERENCES

[1] E. C. van der Meulen. "Three-Terminal Communication Channels". *Adv. Appl. Probab.*,vol. 3, pp. 120–154, 1971.

[2] T. M. Cover and A. A. El Gamal. "Capacity Theorems for the Relay Channel". *IEEE Trans. Inform. Theory*, IT-25(5):572–584, 1979.

[3] P. Gupta and P. R. Kumar. "Towards an Information Theory of Large Networks: An Achievable Rate Region". *IEEE Trans. Inform. Theory*, 49(8):1877–1894, 2003.

[4] L. -L. Xie and P. R. Kumar. "A Network Information Theory for Wireless Communication: Scaling Laws and Optimal Operation". *IEEE Trans. Inform. Theory*, 50(5):748–767, 2004.

[5] L. -L. Xie and P. R. Kumar. "An Achievable Rate for the Multiple-Level Relay Channel". *IEEE Trans. Inform. Theory*, 51(4):1348–1358, 2005.

[6] G. Kramer, M. Gastpar, and P. Gupta. "Capacity Theorems for Wireless Relay Channels". *Proc. 41st Allerton Conf. Communications, Control, and Computing*, pp. 1074–1083, Monticello, IL, 2003.

[7] B. Wang, J. Zhang and A. Host-Madsen. "On the Capacity of MIMO Relay Channels". *IEEE Trans. Inform. Theory*, 51(1):29–43, 2005.

[8] G. Kramer, M. Gastpar, and P. Gupta. “Cooperative Strategies and Capacity Theorems for Relay Networks”. *IEEE Trans. Inform. Theory*, 51(9):3037–3063 , 2005.

[9] M. Gastpar, G. Kramer and P. Gupta. “The Multiple-Relay Channel: Coding and Antenna-Clustering Capacity”. *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Lausanne, Switzerland, 2002, pg. 136.

[10] L. Lifeng, L. Ke and H. El-Gamal. “The Three-Node Wireless Network: Achievable Rates and Cooperation Strategies”. *IEEE Trans. Inform. Theory*, 52(3):805–828, 2006.

[11] C. T. K. Ng, I. Maric, A. J. Goldsmith, S. Shamai and R. D. Yates. “Iterative and One-Shot Conferencing in Relay Channels”. *Proc. IEEE Inform. Theory Workshop (ITW)*, Punta del Este, Uruguay, 2006.

[12] S. C. Draper, B. J. Frey, and F. R. Kschischang. “Interactive Decoding of a Broadcast Message”. *Proc. 41st Allerton Conf.*, Urbana, IL, 2003.

[13] R. Dabora and S. D. Servetto, “Broadcast Channels with Cooperating Decoders”. *IEEE Trans. Inform. Theory*, to appear.

[14] Y. Liang and V. V. Veeravalli. “Cooperative Broadcast Relay Channels”. Submitted to the *IEEE Trans. Inform. Theory*, July 2005.

[15] A. El-Gamal, M. Mohseni and S. Zahedi, “Bounds on Capacity and Minimum Energy-per-Bit for AWGN Relay Channels”. *IEEE Trans. Inform. Theory*, IT-52(4):1545–1561, 2006.

[16] A. Host-Madsen, and J. Zhang. “Capacity Bounds and Power Allocation for Wireless Relay Channels”. *IEEE Trans. Inform. Theory*, IT-51(6):2020–2040, 2006.

[17] J. N. Laneman and G. W. Wornell. “Energy-Efficient Antenna Sharing and Relaying for Wireless Networks”. *Proc. IEEE Wireless Communications and Networking Conference (WCNC)* 2000, vol. 1, pp. 7–12.

[18] X. Bao and J. Li. “Decode-Amplify-Forward (DAF): A New Class of Forwarding Strategy for Wireless Relay Channels”. *Proc. 6th IEEE Workshop on Signal Proc. Adv. in Wireless Comm. (SPAWC)* , New York, 2005, pp. 816–820.

[19] G. Kramer. “Distributed and Layered Codes for Relaying”. *Proc. 39th Asilomar Conf. on Signals, Systems and Computers*, 2005, pp. 1752–1756.

[20] L. Zhixin, V. Stankovic and X. Zixiang. “Wyner-Ziv Coding for the Half-Duplex Relay Channel”. *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Philadelphia, 2005, vol. 5, pp. 1113–1116.

[21] K. Marton. “A Coding Theorem for the Discrete Memoryless Broadcast Channel”. *IEEE Trans. Inform. Theory*, IT-25(3):306–311, 1979.

[22] M. Motani, H.-F. Chong and H. K. Garg. “Backward Decoding Strategies for the Relay Channel”. *MSRI Workshop: Mathematics of Relaying and Cooperation in Communication Networks*, Berkeley, 2006.

[23] R. W. Yeung. *A First Course in Information Theory*. Springer, 2002.

[24] T. M. Cover and J. Thomas. *Elements of Information Theory*. John Wiley and Sons Inc., 1991.

[25] A. Wyner and J. Ziv. “The Rate-Distortion Function for Source Coding with Side Information at the Decoder”. *IEEE Trans. Inform. Theory*, 22(1):1–10, 1976.

[26] F. M. J. Willems. “The Discrete Memoryless Multiple Access Channel with Partially Cooperating Encoders”. *IEEE Trans. Inform. Theory*, 29(3):441–445, 1983.

[27] R. Dabora and S. D. Servetto. “On the Rates for the General Broadcast Channel with Partially Cooperating Receivers”. *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Adelaide, Australia, 2005, pp. 2174–2178.

[28] T. M. Cover. “Comments on Broadcast Channels”. *IEEE Trans. Inform. Theory*, 44(6):2524–2530, 1998.

[29] A. A. El Gamal and E. C. van der Meulen. “A Proof of Marton’s Coding Theorem for the Discrete Memoryless Broadcast Channel”. *IEEE Trans. Inform. Theory*, IT-27(1):120–122, 1981.

[30] A. H. Kaspi. “Two-Way Source Coding with a Fidelity Criterion”. *IEEE Trans. Inform. Theory*, IT-31(6):735–740, 1985.

[31] A. Steiner, A. Sanderovich and S. Shamai. “Broadcast Cooperation Strategies for Two Colocated Users”. Submitted to the *IEEE Trans. Inform. Theory*, August 2007.

[32] R. Dabora and S. D. Servetto. “A Multi-Step Conference for Cooperative Broadcast”. *Proc. IEEE Int. Symp. Inform. Theory (ISIT)*, Seattle, WA, July 2006.